The Computational Analysis of Harmony in Western Art Music

Thesis submitted in partial fulfilment
of the requirements of the University of London
for the Degree of Doctor of Philosophy

Lesley Mearns

July 2013

School of Electronic Engineering and Computer Science,
Queen Mary University of London

I certify that this thesis, and the research to which it refers, are the product of my own work, and that any ideas or quotations from the work of other people, published or otherwise, are fully acknowledged in accordance with the standard referencing practices of the discipline. I acknowledge the helpful guidance and support of my supervisor, Dr. Simon Dixon.

Abstract

This thesis describes research in the computational analysis of harmony in western art music, focussing particularly on improving the accuracy and information-richness of key and chord extraction from digital score data. It is argued that a greater sophistication in automatic harmony analysis is an important contribution to the field of computational musicology.

Initial experiments use hidden Markov models to predict key and modulation from automatically labelled chord sequences. Model parameters are based on heuristically formulated chord and key weightings derived from Schönberg's harmonic theory and the key and chord ratings resulting from perceptual experiments with listeners. The music theory models are shown to outperform the perceptual models both in terms of key accuracy and modelling the precise moment of key change. All of the models perform well enough to generate descriptive data about modulatory frequency, modulatory type and key distance.

A robust method of classifying underlying chord types from elaborated keyboard music is then detailed. The method successfully distinguishes between essential and inessential notes, for example, passing notes and neighbour notes, and combines note classification information with tertian chord potential to measure the harmonic importance of a note. Existing approaches to automatic chord classification are unsuitable for use with complex textures and are restricted to triads and simple sevenths. An important goal is therefore to recognise a much broader set of chords, including complex chord types such as 9ths, 11ths and 13ths. This level of detail is necessary if the methods are to supply sophisticated information about the harmonic techniques of composers. Testing on the first twenty-four preludes of J. S. Bach's Well Tempered Clavier, hand annotated by the author, a state of the art approach achieves 22.1% accuracy; our method achieves 55% accuracy.

no wise fish would go anywhere without a porpoise
(Lewis Carroll)

Acknowledgements

My sincere thanks to: Dr. Simon Dixon, for dedication, wisdom, untiring supervisory support, and the pruning of flowery language; Professor Mark Sandler, for leadership, benevolence, and visionary support of digital creativity; Professor Mark Plumbley, for feedback at the successive stages of my doctorate, fellowship application assistance, and for being a ready source of enthusiasm; the Centre for Digital Music (C4DM), for providing an exciting, stimulating and world renowned research environment.

Thanks to my fellow PhD students: Dr. Matthias Mauch and Dr. Katy Noland for continued encouragement and help; Dr. Rebecca Stewart for being an inspiration; Amelie Anglade for LaTeX wizardry and paper writing advice; and Dr. Emmanouil Benetos for fixing my APA template during a weekend and for always going out of his way to help me. Thanks also to Magdalena Chudy, Steven Hargreaves, Robin Fencott, Yading Song, and everyone else who has provided friendship and support.

Many other people at Queen Mary have given me help and assistance. I would like to thank: the harpsichordist Dr. Dan Tidhar, for friendship, and assistance with the interpretation of scientific papers; Dr. Chris Harte for advising on annotation syntax and giving me his PhD LaTeX template to use; software engineer Chris Cannam, for always making time to help with programming; Kok Ho Huen, for smiley technical support; Melissa Yeo, for fierce reminders which kept me working always; and the staff at Queen Mary's Learning Institute for professional training, especially Dr. Kevin Byron, Dr. Ian Forrestal and Dr. Tracy Bussoli.

Thanks also to: Distinguished Professor Tim Carter of the University of North Carolina for musicology advice; Dr. Phillip Kirlin for friendly thoughts on computational music analysis and the voiced digital scores of Bach; Shelagh and Peter for providing a retreat; my friend Annie Gardiner

for harmonic analysis at the keyboard and innate musicianship; my friends, Tim, Alison and Hilary; and Lily, for quiet canters in leafy woods.

My family have borne with me throughout my studies. Thanks to Grant for taking over the buying of food, to Rita for looking after the kids during ISMIR 2010. Thanks and much love to my children, Jonathan and Jacob, who tolerated oven meals, holiday clubs, and erratic parenting. They invented endless games in the garden whilst I studied.

This doctoral training was financially supported by Engineering and Physical Sciences Research Council (EPSRC) DTA studentship, without which it would not have been possible.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

> What we still don't have is... studies that are grounded in mainstream musicological problems and that make use of computational tools as simply one of the ways you do musicology [Cook, 2005]

The aim of this research is to forge closer integration between the two disciplines of musicology and computer science and to advance the state of the art of music information retrieval (MIR) and computational musicology research. We suggest that a greater level of immersion of musical knowledge into computational approaches to music would be of benefit to both disciplines, potentially creating new directions for musicology research and improving the depth of results produced by MIR applications. The suggestion has been made previously, most notably in Cook's keynote speech at ISMIR 2005 [Cook, 2005] and more recently at IRCAM, detailed in Volk and Honigh [2012]. In this research we are focussing on improving the modelling of harmony, and particularly the underlying chord structure of challenging corpuses which feature a large quantity of inessential notes such as passing notes and neighbour notes. The work described here can be thought of as being broadly applicable to music from the 'common practice' era of western music, a period of time generally accepted as covering the eighteenth and nineteenth centuries. In compositional terms this ranges from J. S. Bach to Debussy approximately.

Musicologists use a wealth of constructs with which to scrutinise music, freely discussing 'harmony', 'rhythm', 'chromaticism', 'tonality', and 'form', as well as more esoteric concepts like 'colour', 'texture' and 'mood'.

Of all of these, the harmonic language of a composer, especially characteristic uses of chords, keys and modulatory technique, dominates music analysis literature. If chord and key information could be captured to a convincing degree using a computer system, opportunities are opened up to yield novel musical insights across a range of musical works, and commonly accepted facets of music history could be substantiated or brought into question based on quantitative data.

Meyer [1973] makes a distinction between *critical analysis* and *style analysis* with respect to the study of music. The former, he describes, belongs in the domain of the musicologist, who concentrates with fine detail on an individual musical work in order to expose its meaning. The latter field of study adopts a broader view of a group of works, with the aim of separating characteristics which distinctively link the group from those which don't.

Inspired by Meyer's writings, Deliege [2007] defines *external* and *internal* similarity relations, where 'external' is the parallel of Meyer's definition of style analysis, and 'internal' refers to repetitions or patterns within a single composition. She proposes the idea of *musical cues*, describing them as 'brief but meaningful structures' which listeners use to either consciously or subconsciously mentally formulate the 'coherence' of a musical work. Deliege states that cues are 'salient elements at the musical surface', and it is this latter term which is fastened upon in Cambouropoulos [2009]. Cambouropoulos points out that current computational approaches to music analysis adopt a definition of 'musical surface' that is the sequence of pitches at the uppermost level of the score, and asserts that in order to be able to obtain deeper and more humanly satisfactory results from computational music processing, current methods must progress beyond this basic understanding, and instead focus on finding ways of transforming the notes into 'complex musical events'.

The aim of this research therefore is to take digital score data and improve methods of key and chord recognition in order to facilitate computational approaches to the analysis of music. Computers seem uniquely suited to the task of rigorously analysing large corpuses, but for computational approaches to be able to deliver results that are deemed to have

musical validity, the kind of constructs implemented must be persuasive, and the nature of their implementation transparent.

1.1 Motivations of Automatic Harmony Analysis

The proposal of this thesis is that a closer intersection between the disciplines of musicology and computer science offers the potential to expedite novel research areas in musicology which exploit the power of computing. Current methods in the field of music informational retrieval tend to constrain experiments to more simplistic features of a musical surface. Building computational models of more complex musical events is thus an important contribution to the field, and may parallel the kind of processing and decision making involved by a human performing the same task. The research presented in this thesis aims to answer the following questions:

1. How can we translate multi-faceted, inter-related music theory concepts into the rigorous representations required by computers?

2. What computational techniques are most effective in refining the accuracy, granularity and level of richness of key and chord information extracted from digital score data?

3. Can a computer program make fine distinctions, such as distinguishing between harmonically essential and inessential notes in complex elaborated repertoire, or subtle oscillations of key, given the ambivalent nature of such musical concepts?

1.2 Contributions of the Thesis

This thesis describes research exploring computational approaches to extracting information about the principles of western musical harmony, including the recognition of complex events such as modulation, and the classification of non-chord tones.

Chapter 2 of the thesis commences with a detailed overview of music theoretic and historic concepts that are concomitant to understanding the

research subsequently presented. The chapter describes core elements of music theory and their relationship to our understanding of musical style, prior to specifying influential philosophical and theoretic work in the field.

Chapter 3 presents a literature review spanning computational approaches to modelling aspects of musical knowledge, recent approaches to distinguishing musical style computationally, and probabilistic and statistical methods in musicology.

Chapter 4 explores the use of hidden Markov models (HMMs) to detect key and key change in automatically obtained chord sequences. The experiments show that key and chord values heuristically derived from music theory produce the most accurate models of key and modulation in comparison to models based upon the results of perceptual experiments. The results of the models are shown to produce harmony analyses that closely match annotations of the same excerpts by Piston [1983]. This research also exhibits the future potential offered by research into audio transcription, by showing that transcribed audio data produces comparable results to similar experiments using symbolic data. The research reveals that although the models yield results good enough to perform style classification experiments based on chord, key and modulation data, the error rate of the automatic chord recognition algorithm is a significant drawback to the broader applicability of the methods. Chord labelling errors resulting from 'non-chord' tones (such as passing notes) are found to impede the production of chord accurate sequences when used in conjunction with music that features melodic movement and decoration. This problem is addressed in the subsequent two chapters of the thesis.

Chapters 5 and 6 recount work to automatically recognise the underlying chords in complex elaborated textures containing a large quantity of non-chord tones. An important pre-requisite step for the research is the production of high quality ground truth data against which the outcomes of the computational methods can be systematically measured. Chapter 5 therefore describes the task of producing hand annotated chord sequences, undertaken by the author as part of this research, for Bach's twenty four preludes in Book One of the Well Tempered Clavier. Hand annotating chord data for complex elaborated keyboard music is demanding on many

levels, consequently the Riemann reference was chosen as the primary source of reference in the production of the data [Riemann, 1890]. The chapter also discusses issues relating to the different theoretical perspectives about Baroque harmony, the paucity of comprehensive sources of reference, the variation in labelling style and choices made in the music reference material that does exist, segmentation issues, and the author's contributions to the improvement of computational annotation syntax in order to accommodate a richer degree of information.

A difficulty with the task described in chapters 5 and 6 is the defining a chord range that provides for broader applicability of the methods, a rich degree of information, and acknowledges the accepted wisdom of music theory. The corpus chosen is highly elaborated and was chosen intentionally because of the challenging nature of the music. Selecting a Baroque corpus however presents issues in and of itself, due to the conflicting opinions between both present day and historical music theorists regarding the chord types that are considered to be valid in corpuses of this period. (Some music theorists contend that no chords beyond simple sevenths are valid.) The issue is somewhat alleviated by the author's close adherence to the Riemann annotations during the production of the ground truth data, in which he labels chords of the 9th and 11th. Examples of complex extended chord types in the music of J. S. Bach from other pre-eminent music theorists are also given. Such issues of music theory are therefore balanced against the goal of the research, to improve computational provision for musicology and music analysis in relation to a broad range of music.

Chapter 6 chronicles the novel chord recognition method used to discover the underlying chords in the test corpus. To be able to access the underlying harmony of ornamental music, the automatic differentiation of chord and non-chord tones, such as melodic passing notes and neighbour notes, is of crucial importance. Due to the fact that digital score data does not always contain voice information, or may be voiced counter-intuitively, a pre-requisite stage of development is to segment the score data into musical voices/linear musical streams. Despite extensive research, automatic voice separation is largely thought to be not completely solvable as the

voiced notation of music is both variable and individualised to composers and editors, and because musical voicing does not adhere to a logical rule set. This chapter explains the novel voice separation method implemented and the evaluation results when compared to ground truth data. The temporal segmentation of the data and a novel method for identifying non-chord tones taking into account both linear position and contextual intervallic relations are described. The methods used to compute musical contour, pedal tone classification, and a measure of metrical and durational emphasis are also related.

The chord method presented in this chapter processes all possible combinations of notes for each temporal segment, as all notes are considered to be potential chord notes prior to processing. Note features are allocated a heuristic value to facilitate the computation of a measure of note importance for each note within the context of the segment. The sum of individual note importance values per distinct note combination is then used to discover the group/groups of maximally salient notes for the segment. The tertian arrangement potential of each note subset is similarly computed, generating the maximally harmonic note combination/s for the segment. The two types of measures, (note features and tertian arrangement), are then compounded to produce a third type of measure, thus producing the most structural, (where structural refers to the concept of notes that are made conspicuous in the musical texture either through articulation or harmonic presence or both), note combination/s for the segment. Intermediate evaluation results are given comparing the notes identified as structurally important to the ground truth data, taking into consideration the notes available within that temporal unit in the score data.

The final stage of the chord recognition method explores the use of weighted and non-weighted chord templates to obtain an optimum chord classification by matching the generated note combination to a chord dictionary. The results are compared to a baseline evaluation, demonstrating a significant improvement over one of the most cited chord recognition techniques [Pardo and Birmingham, 2002]. In addition, by removing inessential tones from the baseline algorithm, this method is also shown

to improve, although not reaching the accuracy of the novel chord method proposed in this thesis.

The final chapter in this thesis contributes a discussion of the work presented and highlights important areas requiring further research in the future.

1.3 Publications by the Author

Publications by the author are listed here:

[Mearns and Dixon, 2010] Lesley Mearns, Simon Dixon. *An Empirical Approach to Musical Style.* In Proceedings of the 3rd International Conference of Students of Systematic Musicology, Cambridge, UK, 2010.

[Mearns et al., 2010] Lesley Mearns, Dan Tidhar, Simon Dixon. *Characterisation of composer style using high-level musical features.* In Proceedings of the 3rd International Workshop on Machine Learning and Music, Florence, 2010.

[Mearns et al., 2011] Lesley Mearns, Emmanouil Benetos, Simon Dixon. *Automatically Detecting Key Modulations in J.S. Bach Chorale Recordings.* In Proceedings of the Sound and Music Computing Conference, Padova, 2011. (Chapter 4)

Lesley Mearns, Simon Dixon. *Detecting Chords in the Ornamental Preludes of J.S. Bach.* (Chapter 6) (under review).

Chapter 2

Music Theory and Concepts

In this chapter components of music theory that underpin the research methods and approaches presented in this thesis are explained. The chapter commences with an overview of polyphony, counterpoint and voice-leading, before moving on to describe harmony, chords, key and modulation and their importance with regard to musical style. This is followed by an explanation of metre and rhythm. The latter parts of the chapter outline some principle formal analytical theories of music that have either influenced, or been symbolised in, the experiments discussed in subsequent chapters, and the chapter concludes with an overview of some of the main data formats used in computational approaches to music.

2.1 Polyphony and Counterpoint

> Music in more than one part, music in many parts, and the style in which all or several of the musical parts move to some extent independently.[Frobenius, 2012]

The term *polyphony* is often used simply to mean music with more than one part, but this is not a strict usage of the word. In the formal study of Western Music, polyphony refers to a style of composition which grew out of the Middle Ages and continued to be refined on into the Fifteenth Century. The central feature of this polyphonic style was the adherence to strict contrapuntal rules primarily formulated in terms of the musical *intervals*, (distance in semitones), formed between successive or simultaneously sounding notes.

Table 2.1: Musical Interval to Semitone Conversion

Musical Interval	Semitones
Unison	0
Minor 2nd	1
Major 2nd	2
Minor 3rd	3
Major 3rd	4
Perfect 4th	5
Augmented 4th	6
Diminished 5th	6
Perfect 5th	7
Minor 6th	8
Major 6th	9
Minor 7th	10
Major 7th	11
Octave	12

Musical intervals have labels which reflect the function of the interval in the context in which they are found. Enharmonic equivalence refers to the notion that an interval may be notated to reflect its role. For example, a tritone (six semitones), is both an augmented fourth and a diminished fifth, varying to fit with linear or melodic voice-leading context (see below 2.1.1). The treatment of musical intervals in a piece of music is strongly characteristic both of historical period and the style of the composer [Piston, 1983]. In the music of Chopin, notated double sharps and double flats are common, for example, Chopin may write a doubly flattened sixth rather than the equivalent, a perfect fifth. A list of musical interval labels and their semitone equivalents are shown in Table 2.1.

The significance of the principles of counterpoint to the development of Western tonal music and beyond cannot be overstated. From the Middle Ages onwards to Bach, Mozart and Beethoven, the independence, shape and movement of the contrapuntal line was of paramount concern, often overriding considerations of tonal relations and harmony. Contrapuntal technique is at the very core of Western musical language largely due to the publication in 1725 of the book *Gradus ad Parnassum*, a pedagogic work written by Johann Joseph Fux (1660 - 1741). The work was hugely

Table 2.2: Species Counterpoint Definitions [Mann, 1943]

Classification	Description
Perfect consonance (PC)	Unison, Perfect 5th, Octave
Imperfect consonance (IC)	Maj/Min 3rd, Maj/Min 6th
Dissonance (D)	Maj/Min 2nd, Perfect 4th, Tritone, Maj/Min 7th
Motion	
Parallel (P+/P–)	Two parts move up or down by the same interval
Similar (S+/S–)	Two parts move in the same direction
Contrary (C>/C<)	One part ascends and the other descends or vice versa
Oblique (Obl+/– Obu+/–)	One part moves up or down, the other remains stationary

influential, studied in depth by master composers including Haydn and Beethoven, as is evidenced by the existence of manuscripts based on Fux' exercises, and copies of the book inscribed by the composers, their teachers or their students [Mann, 1971]. *Gradus ad Parnassum* takes the student through the five species of counterpoint leading to the mastery of four part florid counterpoint. The rules for writing counterpoint are influenced by the compositions of the Italian Renaissance composer Giovanni Periluigi de Palestrina (1525/6 - 1594), and deal with the progression of the contrapuntal line, without "reference to harmony in the sense of organized harmonic progression" [Forte and Gilbert, 1982], p42. The treatment of *consonant* intervals and *dissonant* intervals, also otherwise known as *concords* and *discords* is a central feature. *Consonance*, and its antonym, *dissonance*, can be broadly defined as the psychoacoustic effect of two notes sounded together, the former producing an effect of 'harmoniousness' and the latter, 'roughness' or 'tonal tension' Palisca and Moore [2012]. In species counterpoint new contrapuntal lines are set against a given musical line known as the *cantus firmus*, (literal meaning, 'fixed song'). The cantus firmus, which may be found in any voice but is more often in the tenor, provides the foundations upon which the new composition is built - all other lines are based upon it. The details of species counterpoint definitions and rules are listed in Tables 2.2 and 2.3 respectively, but the crucial characteristic of species counterpoint is that of strong contrapuntal control in accordance with a set of rules.

Table 2.3: Species Counterpoint Rules Mann [1943]

First Species
Note against note: both parts contain notes of same duration. Begin and end on a perfect consonance (PC). Consonant intervals only between the voices. No parallel perfect consonances. From one PC to another PC, proceed in contrary or oblique motion. From a PC to an IC, proceed in any motion. From an IC to a PC, proceed in contrary or oblique motion. From IC to IC, proceed in any motion.
Second Species
Binary metre - two notes against one note of cantus firmus (c.f.). Dissonance allowed to fill in the gap between an ascending or descending consonant skip - i.e. dissonance is a passing note between the interval of a major or minor third. Previous conventions of first species continue to apply.
Third Species
Four notes to each note of c.f. If five crotchets follow each other (asc. or desc.) the first and third must be consonant, the fourth crotchet in the bar may only be dissonant if the fifth note is consonant. All dissonant intervals must resolve downwards by step onto a consonant with the immediately following note.
Fourth Species
Two minims to each semibreve of c.f. but each two minims are on the same pitch and joined by a tie. The first minim must occur on the upbeat, the second on the downbeat. Minims may be consonant or dissonant. Species known as ligature or syncopation.
Fifth Species
Combines all the preceding species rules. The student is advised to write a “singable, melodic line” which contains crotchets, minims and ligatures.

In later style periods, composers such as Debussy (late 19th to early 20th Century), and Messiaen (20th century), invented entirely new sonorities by novel approaches to the combination of musical intervals. A characteristic of Debussy is that of stacking the same interval one on top of the other, creating a totally new sound. Similarly, Messiaen created new musical scales based on new linear intervallic patterns; many of his compositions are based on these ([Neidhöfer, 2005, McFarland, 2005]). Later still, composers such as Penderecki and Ligetti strictly limit the use of the traditionally consonant intervals, and deliberately promote semitone and tritone patterns, so that the music has little or no sense of tonal centre ([Cuciurean, 2000]). Much musicological research is based on uncovering the intervallic patterns that composers employ to create musical works with a particular style, sonority, or structure (e.g. [Antokoletz, 1986, Brown, 2009]). Little of this has been quantified computationally.

2.1.1 Voice-Leading

> The principles of voice-leading always take precedence over considerations of the chord as such, since chords themselves have their origins in the coincidence of melodic movement [Piston, 1983]

It would be rare to read any musicological or music analytical book, article, or paper about a composition without the term *voice-leading* appearing. Sometimes the term is used only a little, in passing, but more often, the voice-leading of a musical work is the central focus of writing aimed at increasing our understanding of music. This is irrespective of the methodological approach used or primary goal of the analysis, as a quick perusal of the titles of articles in music journals will testify. (For examples see [Alegant, 2001, Neidhöfer, 2005, Parks, 1976]).

The term voice-leading is defined as being synonymous with that of part-writing:

> [Part-writing (voice-leading)]. That aspect of counterpoint and polyphony which recognises each part as an individual line

> (or voice), not merely as an element of the resultant harmony; each line must therefore have a melodic shape as well as a rhythmic life of its own. In discussions of part-writing a distinction is made between linear or conjunct motion and movement by leap (i.e. by a 3rd or greater) in a single part, and between various types of relative motion between two or more parts.[Drabkin, 2007]

The concept of part-writing historically precedes that of voice-leading, and tends to be the term used most extensively in analysis of early music (e.g. Palestrina and earlier). Voice-leading, derived from the German, 'Stimmfuhrung', has been dismissed by purists as 'German-American musicological jargon' [McL and Dent, 1950], but one wonders whether the ubiquitous adoption of the term reflects the momentous shift in polyphonic compositional thought said to have occurred around the fifteenth century. Blackburn [1987] cites the change from cantus firmus compositional technique, (successive composition), towards a process of composition in which individual voices are simultaneously composed both in relation to themselves and in relation to the other voices, (simultaneous conception) as "one of the great turning points in the history of music". The article delineates a crucial feature of voice-leading: it has a multidimensional quality, at once embodying the linear dynamic progression of the individual line, and the vertical note formation created by simultaneously sounding notes in other voices. Unlike 'part-writing', the term voice-leading intrinsically infers this quality, a musical concept which is about the simultaneous sounding of individual voices and voices that *lead* somewhere.

2.1.2 Musical Voicing and Compound Melody

Musical voicing is in itself a topic of distinct academic interest due to ambiguity surrounding the concept with respect to instrumental music (e.g. see [Cambouropoulos, 2008]). Although musical voicing originated with part-writing for vocal music, the principles of counterpoint and voice-leading continue to be used as primary compositional techniques in instrumental

music both on the smaller and larger scale [Forte and Gilbert, 1982]. The techniques of musical elaboration used particularly in keyboard music, and the fluctuating musical textures that idiomatic of keyboard music, (between ornamentation and chordal structures), makes it difficult to precisely demarcate the separate musical voices at work at any one moment. In instrumental music it is no longer valid to consider that a musical voice consists of a single note, a single musical voice may contain several notes working together to produce a single linear process. In contrast to this, there is also the concept of compound melody, in which the arrangement of single succession of notes are arranged to give the impression of two or more distinct musical streams rather than a single voice. Compound melodic figuration tends to feature an approximately repeating series of tones and one or more relatively large intervallic leaps, (major sixth or more), which acts as a voice separator. In some cases the two distinct streams may also imply contradictory harmonies or pitch class sets. For example, in Prelude 2 of Bach's Well Tempered Clavier Book One, which can be seen in Figure 5.3 later in this thesis, the notes on the first and third beats of each bar in the upper and lower streams are metrically and registrally accentuated, giving the impression of independent voice processes occurring at a higher structural level. The issue of musical voicing is discussed in chapter 3 with reference to computational work, in which a precise definition of musical voice is required prior to the development of an algorithm.

2.2 Harmony

> Harmony: the study of simultaneous sounds and of how they may be joined with respect to their architectonic, melodic, and rhythmic values and their significance, their weight relative to one another. [Schönberg, 1922]

The "significance of simultaneous sounds"; their "weight relative to one another": Schönberg [1922] gives a definition of harmony that consummately conveys the intricacies of the subject. The term harmony

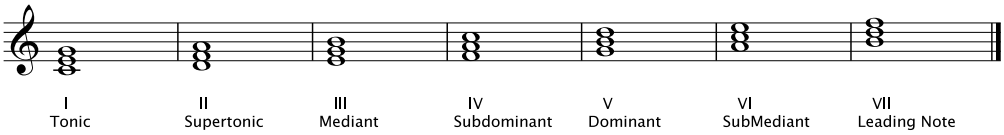


Figure 2.1: C Major Diatonic Triads Schönberg [1922]

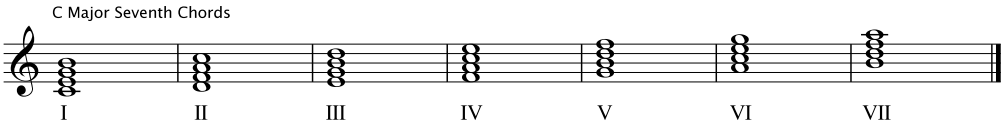


Figure 2.2: C Major Diatonic Sevenths Schönberg [1922]

describes the way in which groups of notes are employed, metrically and melodically, to create harmonic sonorities, the way in which these sonorities are connected, their hierarchical organisation, and the way in which they are used to create larger scale structures and coherent wholes. This chapter commences by defining fundamental precepts, such as chords and keys, before moving onto a discussion of harmonic theory and the relationship between harmony and musical style.

2.2.1 Chords

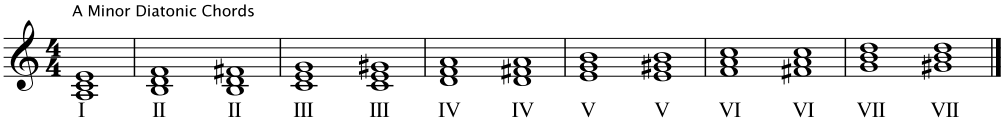


Figure 2.3: A Minor Diatonic Triads Schönberg [1922]

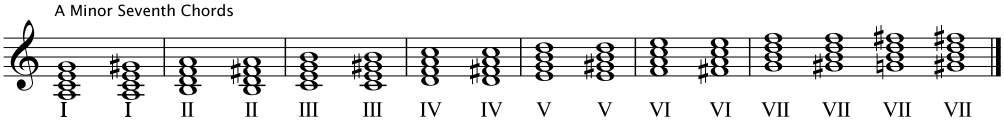


Figure 2.4: A Minor Diatonic Sevenths Schönberg [1922]

Apel [1970] defines a chord as "the simultaneous sounding of three of more tones". The makeup of individual chords can be described by their interval configuration, for example, a major triad in root position is created by placing, above a fundamental note, two further notes, at the interval of a major third (4 semitones), and a perfect fifth (7 semitones). This may also be thought of as the superimposition of tones at the intervals of a major third and then a minor third. The total vertical interval content of a root position major triad is therefore {4, 7, 3}. The minor triad consists of two notes at the intervals of a minor third (3 semitones), and a perfect fifth above the fundamental, a diminished triad consists of a minor third and a diminished fifth above a fundamental, whereas the augmented triad consists of a major third and an augmented fifth above the fundamental.

Triads built upon the degrees of the major or minor scales, generally denoted by the Roman numerals from I to VII, are known as the diatonic triads. In the major scale, the diatonic triads, I, IV and V, are major, II, III, and VI are minor, and chord VII is diminished. Figure 2.1 shows the diatonic triads on the scale degrees of C Major and lists the roman numerals and common names of the scale degrees and Figure 2.3 shows the triads on A Minor.

The concept of chord position and inversion was formulated by theorists in the 17th Century (Lippius, 1612 and Baryphonus 1615), but is predominantly attributed to Rameau due to his endeavours to express the theory of harmony in stricter scientific terms [Dahlhaus, 2007]. Root position, in which the fundamental tone is in the lowest / bass voice, is asserted to be the strongest, most stable chord position. A first inversion (denoted by a b) places the third of the chord in the position of the lowest tone. This arrangement is also known as a 6-3 chord, depicting the different interval quality of the chord - consisting of a third and a sixth above the bass note. The first inversion has a different sound quality to the root position chord; being somewhat less emphatic, it is thought to be less final [Kitson, 1920]. Cadences onto a first inversion are often used mid phrase, to impart a sense of key without affecting the continuation of the phrase. The second inversion (denoted by a c) places the fifth of the triad in the bass voice, also known as a 6-4 because it consists of a fourth

and a sixth above the bass note. This chord position tends to be used in particular situations, for example as a 'cadential 6-4' in which the sixth and fourth are treated as dissonant tones which resolve onto the fifth and third, with a held bass note, of the following chord. Such treatment shows that the 6-4 chord is thought to be unstable.

Chords can also be spaced in open or closed position. A chord in closed position keeps the chord tones close together, with no gaps between which a further chord tone could be placed. An open chord position is as it reads, a positioning of chord tones with some open spaces for further tones in between. Examples of the different chord positions are shown in Figure 2.5.

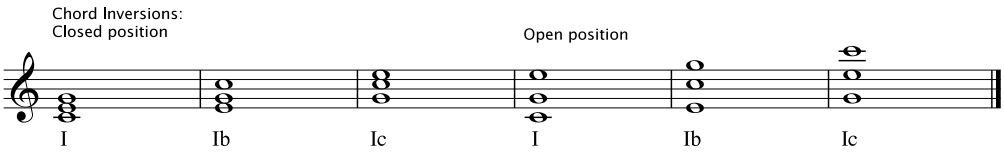


Figure 2.5: The C Major triad in root, first and second inversion, closed and open positions

2.2.2 Doubling, Extended Chords and Tone Omission

Four part harmony, refers to music written for four voices: soprano, alto, tenor and bass, (also known as SATB). Although originating chorally, four part harmony is regularly adhered to as the voice-leading basis of instrumental composition. To meet the requirements of four part harmony, it is necessary to double one of the tones of a triad, thus providing the fourth tone, and conversely in the case of extended chords containing more than four notes, to omit a tone or tones. The omission of tones is a complex area; the aim is to not jeopardise the sound quality of the chord as a result of tone omission but to allow a representation of the chord to fit with melodic movement. The rules about which tones to omit are not conclusive and the tones omitted in music practice tend to vary in accordance with voice-leading and musical context. Kitson [1920] states that the triadic tone most usually doubled is the fundamental note because

the third becomes too prominent in the sound mix if doubled, whereas the 6-4 chord doubles the fifth, thus giving it a unique sound quality.

From the basic triads many different chords can be created by the superimposing further notes. The most common addition is that of the seventh, the tone a third above the fifth of a triad. Piston [1983] defines the seven interval configurations of seventh chords in root position and says that seventh chords have their own rules with regard to treatment, doubling and tonal makeup. (Consider all possible chords made up by the different patterns of successive intervals from 3-3-3 to 4-4-4, shown in Table 5.2 in Section 5.6). The most common seventh is the dominant 7th, in which the minor 7th scale degree is superimposed onto a chord V. Piston [1983] asserts that the dominant 7th strongly implies resolution onto the tonic chord and unequivocally implies either the major or minor key of the tonic note.

Theorists differentiate between *consonant* (stable) and *dissonant* (unstable) chords on the basis of their intervallic content: a consonant chord contains consonant intervals (please refer to section 2.1 with the proviso that the intervals classed by musicians to be consonant / dissonant have mutated during the course of history); a dissonant chord contains at least one dissonant interval (e.g. second/seventh/tritone). Hindemith [1942] considers that tritone content is an overriding harmonic factor in the musical treatment of a chord, overriding the importance of notes such as the root. He divides chords into two groups; those containing tritones, and those not containing tritones.

Extended, or complex chords, (sevenths, ninths, elevenths, and thirteenths) are all firmly placed in the category of dissonant chords. There is a great deal of literature about their definition, usage, treatment and stylistic implications. The so called 'dominant ninth' (Piston [1983]) is where the diatonic ninth from the root is added to the V7 chord, and the 'dominant thirteenth' is often used as a displacement of the fifth note in the chord V which resolves onto this note. (From a computational perspective this kind of arrangement is easily misinterpreted as a first inversion chord on the thirteenth note). Figure 2.6 shows some extended chords in common arrangement in four part harmony.

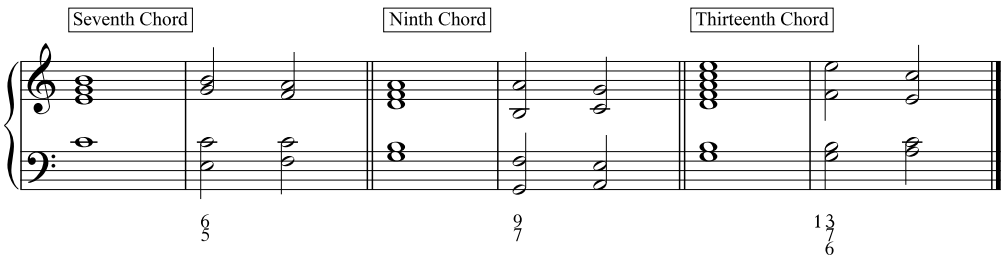


Figure 2.6: Example Arrangements of Extended Chords

Another widely accepted concept in relation to chords and harmony is that of *chromaticism*, most easily understood as the introduction of tones foreign to the diatonic tones of the scale. In common practice repertoire this most often occurs in the form of semitonal voice-leading movement for the purposes of musical colour and interest [Kitson, 1920]. Chromaticism is associated with particular composers, for example, Chopin [Kramer, 2012].

In homophonic works, block chordal structures are the primary chord type. Complex polyphony is also acknowledged to elaborate chordal structures, with the works of J. S. Bach being some of the most well known compositions embodying both the principles of counterpoint and chordal elaboration [Kirkpatrick, 1984]. Ledbetter [2002] discusses the influence of the writings of music theorist Frederich Erhard Niedt (1674-1708), whose work *Musicalische Handleitung* details techniques for melodically embellishing a figured bass structure. Kitson [1907] comments that even for composers as early as Palestrina, the ability to decorate a basic tonal structure is seen as the skill which renders master composers more prominent in comparison to lesser composers,.

2.2.3 Cadences, cadential structures

A cadence, also termed close, is the use of two chords in succession to demarcate a boundary, such as the end of a phrase. The cadence which is most final, and which firmly defines a key in root position, is the 'Perfect Cadence', which consists of a V - I or V7 - I progression. Other cadential

types include the 'Plagal Cadence', IV-I, and 'Interrupted', V - VI. This latter introduces an important aspect of cadences, which is that a sense of expectation has been created as to how the pre-emptive chord will resolve. A composer can use this expectation to create an element of surprise, by resolving the chord in an unexpected way, or by not resolving the chord at all but moving on to a further dissonance. Composers often use cadential sequences in the lead up to a final cadence, maintaining an element of suspense and thus creating a greater feeling of resolution when the final close arrives. For a detailed discussion of cadences refer to [Piston, 1983].

2.2.4 Figured Bass

Figured bass is the principle of providing musical interval figures indicating the expected or required harmonic configuration of notes above a bass part, or with which to fully harmonise a soprano and a bass line. A root position triad is a 5-3, i.e. a fifth and a third above the bass note. A first inversion triad is a 6-3, and a second inversion triad a 6-4. By convention a 5-3 is not notated, the 6-3 is abbreviated to 6, a 7-5-3 (a root position seventh chord) to 7 and a 6-4-2 (a third inversion seventh) to 2. Figured bass was primarily a practical guide for the performing musician in the 17th and 18th centuries. It continues to be used to this day to denote the harmonies of music from these historical periods by music theorists.

2.2.5 Key and Modulation

An important concept in musical composition is that of key change, also known as modulation. Closely related is the idea of key relationships and key distance: keys which are separated by just one or two additions or subtractions of accidentals in the key signature are thought to be more closely related to one another than keys with quite different key signatures. The concept can be visualised by studying the circle of 5ths, shown in Figure 2.7. Schönberg defines a set of key circles encapsulating the hierarchy of key relationships, listed in Table 2.4. The final circle, number 7, contains keys which are the enharmonic equivalent of keys listed in previous circles.

Change of key within a composition is often brought about by a chord

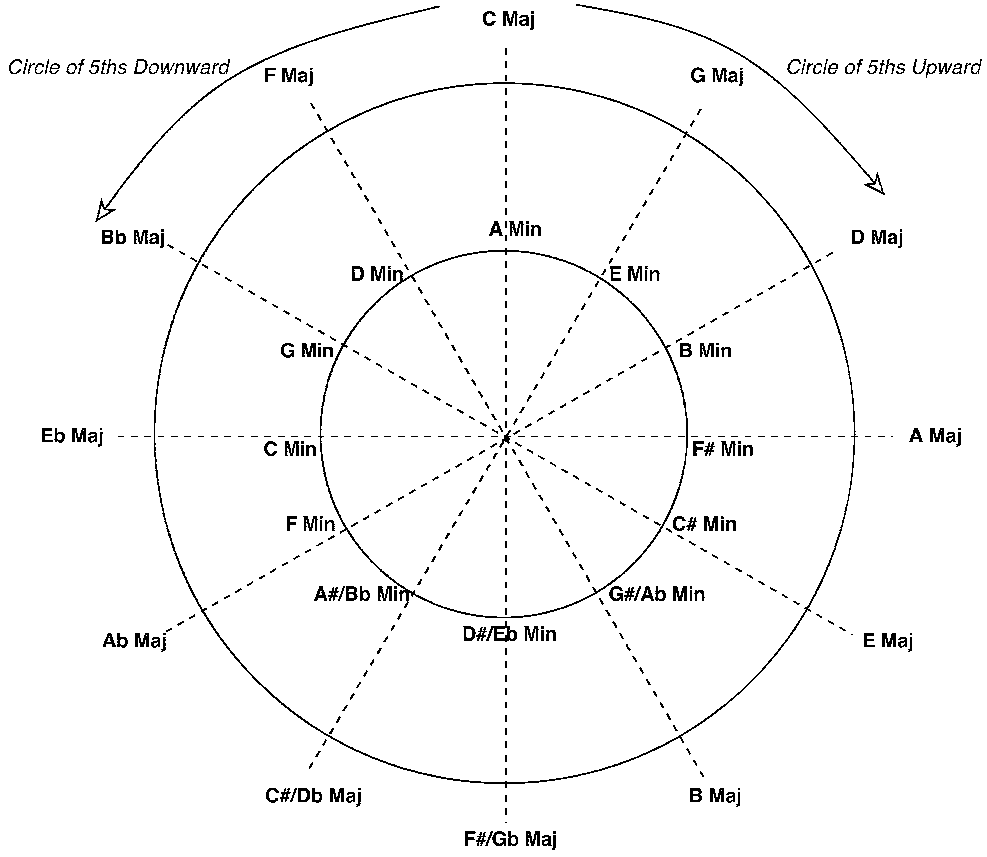


Figure 2.7: The Circle of 5ths

which is common to both the initial key and the new key, commonly referred to as a pivot chord. In practice there may often be a series of pivot chords together which have a dual key function until the move to the new key becomes apparent. This type of modulation is known as a diatonic key change. An alternative method is chromatic modulation, made via the introduction of a chromatic chord that is used as a means of establishing a new key area. An enharmonic modulation exploits the enharmonic reinterpretation of tones within a chord by resolving the enharmonic interpretation of the notes. This method facilitates key changes between distantly related keys.

How key changes are effected within a composition, the frequency and symmetry of key change, the keys moved to, and the use of a key structure

Table 2.4: Key Circles for C Major / A Minor [Schönberg, 1922]

Key Circle	Key 1	Key 2	Key 3	Key 4
First Circle:	G Maj	E min	F Maj	D min
Second Circle:	D Maj	B min	B♭ Maj	G min
Third Circle:	A Maj	F♯ min	E♭ Maj	C min
Fourth Circle:	E Maj	C♯ min	A♭ Maj	F min
Fifth Circle:	B Maj	G♯ min	Db Maj	B♭ min
Sixth Circle:	F♯ Maj	D♯ min	Gb Maj	E♭ min
Seventh Circle:	C♯ Maj	A♯ min	Cb Maj	A♭ min

of delineate the architectural boundaries of a composition, are considered to be important style features of individual composers, groups of stylistically similar composers, and historical period, begetting a vast array of musicological literature. Ferris [2000], for example, discusses the extraordinary modulatory technique of C. P. E. Bach, Bribitzer-Stull [2006] traces the origins of nineteenth-century composers' uses of chromatic major-third relations, and Schachter [1987] revisits the role of modulation in relation to composers of different style periods.

2.2.6 Harmonic Theory

The use of harmony has evolved and mutated during the course of history, from the earliest harmonic inceptions of the Middle Ages, to the vast range of vertical formations used in modern composition. The sheer wealth of theoretical literature and discourse, commencing with the theoretical writings of the 15th Century, to contemporary publications, gives testimony to the complexity, diversity of thought, and variation of viewpoint on the subject.

Dahlhaus [2007] describes, that following the early theories of Zarlino in which harmonic relationships are related directly to string length ratios, one of the most influential theorists is Jean-Philippe Rameau (1683-1764). Rameau's systematic rationalisation of chords and harmonic progressions evolves and mutates over several decades of influential theoretical work commencing with the *Traité de l'harmonie* of 1722 and culminating in the

Nouvelles Réflexions of 1760. His conception of a fundamental bass and reconciliation of 9th and 11th chords in terms of 7th chords with additional roots continues to dominate views on music theory today (e.g. [Gosman, 2000]). Similarly authoritative is Riemann's theory of functional harmony, in which chords are related to their tonic, subdominant and dominant functions. Chords are not treated as entities in themselves but in relation to their function or role within the bar, phrase, movement or musical work. The diatonic chords are labelled with Roman numeral notation in relation to the scale degree that constitutes the chord's fundamental note. A single chord may have several possible functions or labels, depending on its key context. For example, an F Major triad is a chord IV in C Major, a chord I in F major, a, chord V in Bb Major, and so on. To be able to allocate a Roman numeral label to a chord, the key context of the chord must be known. In order to understand harmony, Dahlhaus [2007] states, one must be able to relate chord content, already understood in terms of its role within a key, to metre, musical phrasing, and form. Importantly he expresses the idea that harmony is as much a structural principle in ancient and medieval music as it is in later tonal eras.

2.2.7 Harmonic Analysis

The analysis of harmony in a composition is not an exact science. Non-chord tones, unusual rhythmic emphasis, and chromatic elements, are exploited by composers to create ambiguity and areas of harmonic tension. Harmony analysis is subjective; although analysts will agree for the most part about chord and key designations when the harmony is relatively transparent and clear, opinions diverge in situations of ambiguity about chord function, key, tonal centre, and about which notes are structurally more important than others, and why. (Consider for example the analytical difference of opinion regarding the theme of the first movement of Mozarts K331 - is the C♯ more important than the E or vice-versa?) The variation of opinion between two people is commonly cited as a problem in the field of Music Information Retrieval where unequivocal ground truth data is required for system evaluation (e.g. see [Smith et al., 2011], and

is supported in work by Krumhansl; her resolution to the problem is to obtain ground truth data about key strengths in Bach's C Minor Prelude (Book II) from two musical experts rather than one [Krumhansl, 1995]. In relation to the research in this thesis the topic is discussed in more depth in section 5.5). There are also different ways of going about the analysis of harmony. One method is to designate every single chord label in relation to the overall key of the piece, rather than in relation to the local key context. Piston [1983] terms this 'literal roots'; for a thorough review refer to his chapter on harmonic analysis. The advantage is that one cannot really argue about the chord label. The disadvantage is that it can lead to some rather obscure labelling, and perhaps more significantly, does not acknowledge the key processes at work within a piece, nor even, it could be argued, the function of the chords.

A more usual approach is to label chords in relation to their role within the local key context. This method is the preferred method if one wishes to develop a more detailed understanding of musical form, such as sonata form, which is primarily defined by its higher level harmonic structures. The method thus requires the analyst to discover the key relationships being used to create coherence in a work. To perform this kind of analysis, a certain level of musical skill is needed. One must be clear (at least in one's own mind) about the definition of modulation and chromaticism, and how keys and chords interact to create harmonic structure. At what point does the use of chromatic chords become modulation, or at one point is a seeming modulation just chromaticism? The analyst needs to be able to judge how strong the modulation is, for example, whether it is just briefly passed through, whether it is expressed weakly or strongly, with inverted or root position chords. It may be a modulatory sequence with a long term harmonic goal. Importantly, many musicologists do not consider a modulation to be a modulation unless there is a cadence in the new key [Kitson, 1920].

A further method is to analyse music in the light of 'secondary dominants'. The primary function of the secondary dominant, as detailed in [Piston, 1983] pages 282 to 284, is that the purpose of a new key is to emphasise the strength of the tonic key via tonicisation of dominant or

subdominant harmony. Although the new key area is heard as a modulation, it is thought to be *felt* in relation to the home key [Wishart, 1956]. The notation of secondary dominants takes the form of, 'IV of IV', or 'IV of V' where the first number refers to the chord, the second to the key, identifying the relationship of both chord and new key to the original tonic. The extent to which a key is actually perceived as a secondary dominant has been brought into question by perceptual experiments with listeners [West Marvin and Brinkman, 1999].

2.3 Harmony, Counterpoint and Musical Style

The use of harmonic relationships to carve out a sense of formal balance and symmetrical proportion is seen by musicologists as the central characteristic of the 'Classical style', the musical style which emerged during the second half of the eighteenth century [Mellors, 1957]. The development of a new harmonic language is regarded as the catalyst which induced a period of new creativity and expression to emerge from the contrapuntal control of the Baroque era [Rosen, 1971]. It is seen as a crucial turning point in the history of music, marking a departure from old traditions, and the beginning of a new musical era [Piston, 1983] The progress of harmony throughout history can be charted through the works of the master composers of western music: commencing with the early styles of Palestrina, through the works of Monteverdi and Bach, to Haydn, Mozart, Beethoven; and on to the composers who stretched the definition of harmony further and further with increasing levels of chromaticism (Brahms, Debussy, Ravel, Mahler, Wagner), to polytonalality (Stravinsky, Britten, Bartok), and on until the eventual deterioration of harmony resulting in atonality and serialism (Schönberg, Berg, Webern) [Piston, 1983, Mellors, 1957, Grout, 1980]. Many recent composers have returned to the basic building blocks of harmony within a new compositional framework such as minimalism (Glass, Reich, Nyman, Pärt) whereas others have innovated with dense counterpoint and polyrhythm to create novel sonorities (Penderecki, Ligeti).

The role of harmony, in conjunction with the principles of counterpoint

outlined earlier, are clearly understood in musicology to occupy an important place in the definition of musical style. It is the chord qualities, the way in which they are used to create a sense of tonality, of consonance and dissonance, and an expectation of leading somewhere, and the way in which harmonic structures are expressed by contrapuntal processes, that are the primary distinguishing factors of musical style in western classical music. This dualism between the structural-vertical (harmonic) and linear-horizontal (melodic) dimensions of music is one of the most difficult aspects of music to capture computationally. A computer system tends to focus on a single dimension at a time; simultaneously occurring multidimensional processes are particularly difficult to model.

2.4 Metre

This section outlines those aspects of metre and beat that are a necessary adjunct to understanding research presented subsequently in this thesis. The details and explanation of the concepts presented here are indebted to [London, 2007] who gives a detailed overview of the topic. The upper portion of a time signature defines the number of metrical units per bar. The lower portion defines the unit of measurement: 2 corresponds to minim, 4 to crotchet, 8 to quaver, 16 to semiquaver, and so on. Meters with an upper time signature value of 2, 3, or 4 are known as *simple meters* (for example, $\frac{2}{4}$, $\frac{3}{2}$, $\frac{4}{8}$), whereas *compound meters* multiply the number of units in simple meters by 3 (for example, $\frac{9}{8}$, $\frac{12}{4}$, $\frac{6}{2}$). The concept of metre implies that there is a perceivable *beat*, i.e. a rhythmic pulse, best defined as that to which one could tap one's foot in time to a piece of music. Metres are classified as being *duple*, *triple* or *quadruple* respectively, according to whether there are 2, 3, or 4 *beats* per bar. The durational value of a *beat* does not necessarily coincide with the unit of measurement specified by the lower part of the time signature; a beat can consist of a combination of these units. For example, the time signature $\frac{4}{4}$, is *simple quadruple*, and has 4 beats of crotchet value per bar. In contrast, the time signature $\frac{6}{8}$ is *compound duple*, and consists of two beats per bar of three quavers

Table 2.5: Time Signature Types Lookup Table.

Time Signature	Signature Type	Beat Groupings	Shorthand
2/2 2/4 2/8	Simple	Duple	S/d
3/2 3/4 3/8	Simple	Triple	S/t
4/2 4/4 4/8	Simple	Quadruple	S/q
6/8 6/4	Compound	Duple	C/d
9/2 9/4 9/8 9/16	Compound	Triple	C/t
12/8 12/16	Compound	Quadruple	C/q
24/16	Compound	Octuple	C/o

each. Similarly, the *compound quadruple* time signature $\frac{12}{16}$, has four compound beats of three sixteenth notes per bar. A range of commonly used time signatures and their classifications is shown in Table 2.5. Irregular time signatures are possible and feature in twentieth century composition especially.

2.5 Formal Analytical Theories of Music

A brief overview of formal analytical theories of music is included here for the information of readers wishing to pursue similar lines of research in the future. The theories are related in order of importance with respect to their influence on this work, with pitch class set theory having only a slight bearing on subsequent work, simply the representation of groups of pitches by pitch class sets.

2.5.1 Schenkerian Theory

A description of Schenkerian analysis is included in this section of the thesis because the concept of an 'underlying structure' in music is central to the approach to capturing chords detailed in chapter 6. Heinrich Schenker (1868-1935) propounded an analytical approach to tonal music based on the idea that the surface detail of a tonal musical work is a 'prolongation' of a particular note, chord or harmony, and that all musical works of the tonal era can be 'reduced' down to an underlying fundamental structure via a series of hierarchical levels [Schenker, 1979]. The notion

of underlying structure is explained visually in the article by Heidi Siegel, in which excerpts of Schenker's intellectual correspondence with the artist Victor Hammer are quoted [Siegel, 2006]. In their correspondence, Hammer equates his diagram of the visual structure of his portrait of a patron, Parmenia Ekstrom, to Schenker's concept of 'underlying structure' in music, stating that this is the 'purely visual structure that is not seen but apprehended' in his paintings (page 89). In their letters Schenker also refers to the idea of 'musical space', particularly *Quint-Raum* (space of a fifth) and *Terz-Raum* (space of a third), which he proposes are the fundamental building blocks of music that are filled by the 'unfolding *Urlinie*' or structurally descending scale progression (page 92). In Schenker's view, all works are based upon an *Urlinie* with a supporting harmonic structure known as the *Ursatz*, but there have been criticisms of his published analyses that the reductions were biased in favour of this [Salzer, 1982]. Schenkerian theory is intricate and complex and the reader is referred to [Pankhurst, 2008] for a more detailed understanding of the subject matter. Nonetheless, internationally and throughout history, no other musical theory has rivalled that of Schenker in terms of influence or the adoption of its concepts. The theory has excited extensive and wide ranging academic discourse from many different perspectives [Cook, 1989, Littlefield and Neumeyer, 1992]. The theory itself has reached far beyond the tonal music it was originally devised to explain, with transmutations of the theory extending into analyses of post-tonal composition (e.g. [Lerdahl, 1989] and spawning new theoretic approaches to music (e.g. [Lerdahl, 2001]).

Computational approaches to Schenkerian analysis include the work of Phillip B. Kirlin and Paul E. Utgoff [Kirlin and Utgoff, 2008], and the more extensive researches of Alan Marsden (see for example [Marsden, 2011, 2010, 2007]). Marsden reports that one of the biggest challenges facing a computational approach, aside from the inherent subjectivity of this type of analysis, is the sheer quantity and complexity of data that must be processed in order to achieve a Schenkerian reduction. The effect of this is that the computer is only able to process short excerpts [Marsden, 2010].

2.5.2 A Generative Theory of Tonal Music

Lerdahl and Jackendoff [1983] made a notable contribution to the field of music theory with 'A Generative Theory of Tonal Music' (GTTM). It is based upon Chomsky's linguistic methodology combined with Schenkerian music theory. In the opening paragraph of Chapter 1, page 1, the author's describe the goal of the work to be:

> a formal description of the musical intuitions of a listener who is experienced in a musical idiom.

The approach taken is to arrive at a set of well formedness rules to obtain the definition of a group, and a more subjective set of preference rules that take into account perceptual preferences, in the domains of Metre, Grouping, Time-Span, and Prolongation. The influence of linguistic theory is evidenced in the use of tree like structures to explain the structure of pieces of music, however they seem a little remotely related to the music they represent.

Despite being a theory about *tonal* music, the book notably omits any detailed explanation of tonal relations or thematic/motivic processes in music. Hamanaka et al. [2007], have implemented aspects of the theory computationally, but state that the original theory is not directly adaptable to computation due to a lack of direction about how to proceed when there are conflicts between rules. The applicability of the GTTM to real world problems such as transcription, or style recognition has not been fully tested, consequently there is scope for further research into the application of a musical grammar of this kind. GTTM theory is used to inform the implementation of measures capturing the metrical strength of notes in the work presented in chapter 6.

2.5.3 Narmour's Melodic Implication/Realisation Theory

Narmour [1992] places melodic processes in the domain of cognitive psychology in his extensive theory of melody, 'The Analysis and Cognition of Basic Melodic Structures: The Implication-Realization Model'. Narmour defines a theory of 'process', in which small melodic intervals imply

melodic continuation, and 'reversal', in which large intervals imply reversal or difference. Similarity and difference are judged according to melodic context and this appears to be interpretive. Narmour also defines whether a tone has a structural or nonstructural role, depending upon whether the tone concludes, or fails to conclude, a sequence of preceding tones which are implicative in nature. The theory has motivated quantitative experimentation testing the theory e.g. [Krumhansl, 1995].

2.5.4 Pitch Class Set Theory

Forte [1973] developed pitch class set theory to expedite the analysis of atonal and serial music; i.e. music that intentionally avoids any notion of having a key or tonal centre. A pitch class set is any group of simultaneously occurring pitches, although it can also refer to a group or combination of pitches musically expressed in linear or diagonal formations. A pitch class set does not have the concept of reinforcing a sense of key, and crucially differs from tonal harmony in that all pitches are included. To define a particular pitch class set, the constituent pitches are translated into a unique set of pitch classes, commencing with pitch class 0 (C), progressing chromatically up the scale to 11, (B). The normal order of a pitch class set is when the pitches are arranged in ascending order within a single octave and then cycled around until the first and last pitches have the smallest interval between them, for example, the set [0,3,2] in normal order is [0,1,2] taking into account transpositional and inversional equivalence. Central to the theory are methods of ascertaining whether pitch class sets are equivalent by transposition or inversion: for example, an interval of a perfect fourth is equivalent to that of a perfect fifth, an equivalence which would not be accepted by tonal composers such as Bach. An important relationship between two pitch class sets is the 'interval vector', which departs from describing the pitch content of a set and looks instead at the interval content. Two pitch class sets are said to be related if they share the same interval vector, even if their pitch class set classification is different. Forte [1973] gives a complete list of pitch class sets, their names, pitches and interval vectors.

Chapter 3

Literature Survey

This chapter presents the research to date relevant to the work described in this thesis, particularly methods of modelling music and formulating representations of musical constructs using computational methods. Detailed attention is given to work that has a direct impact on the research presented later in this thesis.

3.1 Counterpoint and Voice-leading

The musicologist Robert D. Morris [1] has published extensively about music theory including a detailed and thorough account of counterpoint and voice-leading. Morris [1998] formally describes features of counterpoint before going on to discuss the theory of using a 'Tonnetz', in which pitch classes are represented in a two dimensional pitch class array to represent transformations between pitches. His ideas are influenced by previous work done by David Lewin, described below. Robert Morris's systematic definition of counterpoint features are valuable to a computational implementation, however the simultaneous representation of structural-vertical aspects of music and linear-horizontal aspects of music that humans perceive so readily is peculiarly difficult to encode in a computer system. Much of the computational research done so far has therefore been limited to simpler one-dimensional representations of voice crossing or voice separation, rather than voice-leading. The idea of voice-leading analytical

[1] http://www.esm.rochester.edu/faculty/morris_robert/

layers is central to that of Schenkerian analysis [Pankhurst, 2008], consequently work done on automatic Schenkerian reductions needs to take some elements of voice-leading into account. Kirlin [2009] identifies a type of voice-leading layer that can be identified relatively easily by a computer - that of the linear-horizontal stepwise progression occurring at a higher structural level such as the first beat of every bar. Diagonal and vertical inter-voice relationships are not accounted for.

Laurson et al. [2008] describe a similar problem in their work on the visualisation of computer assisted music analysis (PWGL), saying that 'voice-leading rules tend to be harder to formulate than melodic and harmonic rules as they deal with both melodic and harmonic formations at the same time'. Their system consequently limits the representation of voice-leading concepts to that of recognising when voice-crossing takes place in a musical work. The musical voices in question are restricted to one note per voice - a topic discussed further below.

David Lewin's use of vectors to represent voice-leading progressions between two pitch class sets successfully represents the interval crossing relations of two pitch class sets [Lewin, 1998, 2001]. His theory exploits pitch class set theory and is primarily applicable to atonal and serial compositions by composers such as Schönberg, but it also holds promise for the processing of voice-leading in tonal music. An issue relating to voice-leading functions in tonal music is the interval equivalence expressed. (See section 2.5.4 on pitch class set theory). Lewin's voice-leading functions take no account of register, and therefore although one of his functions claims to account for 'total potential voice-leading', in practice it does not do so. Lewin's work is of interest due to his detailed research into the nature of transformations from one pitch-class set to another and how this relates to music theory, rather then concentrating on the actual pitch content of the pitch class set. He also pays attention to the intervallic content of music, which as he points out in the first paper cited, is not necessarily a secondary feature of music, but in some cases a primary one. He gives as an example a sequence containing very individualistic intervals in George Crumb's *Makrokosmos* for piano: 'the constancy of those numeric values, from each stage of the progression to the next, is a feature in

its own right'. The application of Lewin's theories in methods of harmony extraction in tonal music remain to be researched.

S. T. Madsen details experimentation with an evolutionary algorithm to model species counterpoint in which randomly generated sets of notes are evaluated in accordance with encoded species rules [Madsen, 2005]. The sets of notes that give the best fit are kept and slightly mutated and the process continues until a set of notes is achieved adhering to the defined rules. The work is based on three cantus firmus parts and explores a systematic application of species counterpoint rules, however musical knowledge is not enlisted to either assess or improve the musical quality of the result, and he reports the results to be disappointing from a musical perspective. Using the same cantus firmus used by Madsen, Eduardo and Roberto Morales explore the application of Inductive Logic Programming (ILP) to learn first species counterpoint rules [Morales and Morales, 1995]. They successfully generate a second musical part in note against note style against the cantus firmus, and they express an aim to incorporate these learned rules in a compositional system. Mearns et al. [2010] uses machine learning tools to attribute digital scores to individual composers based on musical features extracted about use of species counterpoint rules, consonance and dissonance levels, vertical intervals and tonality in two corpuses. A composer classification task is performed to test the ability of the feature sets to discriminate composers, yielding moderate accuracy. The experiments show that abstracting meaningful information is challenging, however the results give promise for practical applications such as style recognition and music recommendation. It is surmised that by improving the capability of the program to abstract high level musical constructs, the classification results and the insights given by the results will also improve.

All of the above implementations feature similar problems with respect to capturing contrapuntal processes in music. The problem relates to the ways in which a computer system is able to store information, for example, in the form of single dimensional storage types which do not easily adapt to a system of references to track inter-relationships of notes. The principles of music analysis are loose, consequently there is an interpretive and

subjective step involved in encoding the relative structural importance of notes, and any implementation will require simplifying assumptions to be made.

3.2 Modelling Musical Voices.

The separation of MIDI data into musical 'voices', which resemble either the way that individual musical lines are written out in the score, or the human perception of auditory streams, is also a challenging problem. One part of the problem is that the notion of voicing has more than one definition (please see section 2.1.1). The majority of computational approaches to voice separation adopt a standard understanding of the term 'voice', which is that a musical voice is a monophonic series of successive notes which do not overlap in time. Importantly, in this definition there cannot be more than one note per voice at any one time. The definition is closest to that used in early vocal styles, in which each 'voice' corresponded to a single part, a definition which is the most convenient for a computer implementation. Nonetheless as mentioned in preceding sections, the restriction of one note per voice does not hold true for keyboard music. As early as Bach's *Well Tempered Clavier*, we see fluctuating numbers of concurrent notes; the musical voicing is not consistent throughout. Varying musical textures are an important feature of the idiom, and the idea of a single note per voice is flawed.

The majority of computational algorithms developed with the aim of deriving a set of musical 'streams' or 'voices' from MIDI data use roughly similar sets of criteria for allocating notes to voices. The principles are inspired by studies of human perception of auditory streaming, and in particular how music is integrated or segregated into streams [Bregman, 1990]. There are two principles which prevail in computational implementations; the first is temporal continuity, i.e. the sequence of notes must be contiguous, and the second is pitch proximity, which is the idea that notes which are closest in pitch are more likely to be perceptually grouped into a single voice. Most algorithms have a limitation of a single note per voice; approaches to the maximum number of voices per score vary, from

manual input to being equivalent to the size of the largest chord. (For example see [Kilian and Hoos, 2002, Kirlin and Utgoff, 2005, Madsen and Widmer, 2006, Temperley, 2001].)

It has been suggested that the problem is not fully solvable with the amount of information available to the system [Marsden, 1992, Kilian and Hoos, 2002]. Marsden [1992] uses J. S. Bachs Fugue in G♯ Minor from Book I of The Well Tempered Clavier to take us through the intractability of the problem. His first model, which defines a rule 'closest', whereby notes are conjoined into a single voice based on the end of one note coinciding with the beginning of the next note, and the next note being closest to it in pitch. This first model is only successful until the beginning of the fourth bar, where the note in the same voice as notated in the score is in fact not the closest to it in pitch, thus it is incorrectly linked to the lower part. Marsden goes on to build a series of models to try to deal with the problem. One of the difficulties is that in attempting to address the various ways in which Bach's fugue digresses from consistent and logical 'rules', is that the design of the model becomes overfitted and loses generality. Moreover, the introduction of rules to resolve more complex voicing situations then results in incorrect voicing in the previous cases where simpler models had selected the correct allocation.

Madsen and Widmer [2006] present an approach to separating voices in MIDI which is inspired by Temperley's well-formedness rules [Temperley, 2001], including the restriction of one note per voice and the creation of a contiguous sequence of note events per voice. They use a cost function to ensure that certain preferences are adhered to when assigning notes to voices, in particular, minimising leaps between notes in all voices, minimising the number of voices, and minimising the number of rests within a voice. Their approach allows the dynamic creation and termination of a voice, and also, in contrast to other approaches, allows voices to cross, although voice crossing incurs a higher cost. They evaluate their results using Bach's three part inventions and fugues and report that the principle of pitch proximity is insufficient to solve the problem. Initial experiments with pattern matching to improve results show only a small improvement but they conclude that pattern matching may be able to further improve

their algorithm's performance.

Cambouropoulos [2008] directly tackles the standard understanding of musical 'voice' and computational models which make an assumption of a single note per voice. He questions the reasoning behind David Huron's work on tone and voice [Huron, 2001], in which Huron asserts that the purpose of voice-leading is to create 'perceptually independent musical lines', linking this in with the perception of auditory streams. He points out that Huron's article neither defines what a musical voice is, nor makes explicit the nature of the relationship between a voice and an auditory stream. In addition, the assumption of only one note per musical voice results in the discounting of a large proportion of music. Cambouropoulos argues that any musical voice related algorithms must be able to accommodate more than one note per voice to be truly functional in a computer system. His *Visa* algorithm groups synchronous notes into a single stream and notes that overlap and are not synchronous are placed in different streams, thus a series of chords would be grouped into a single voice. The method is particularly suited to keyboard music in which there may be a set of chords in the left hand and a melodic upper part, which would constitute a different voice, and is much more in keeping with the style of many keyboard compositions. The algorithm in its current state is unsuitable for use with complex contrapuntal works such as the *Well Tempered Clavier* in which compound melody is frequently a feature (see section 2.1.1). Prelude 2, the *C Minor prelude*, for example, would be grouped into one single stream because it consists of synchronous semi-quavers, rather than the two notated streams or possibly the more desired result - four musical voices. The concept of musical parallelism as implemented in *Visa* therefore requires further adaptation in order to detect compound melodic streams, for example, parallel intervallic movement as well as metrical synchronicity. This type of modification is not straightforward due to the fact that parallel intervallic movement does not necessarily consist of progressions of exact intervals. Further work is needed to thoroughly explore the topic (please see future work).

Other voice-leading related studies by Cambouropoulos include his paper about 'Auditory Streams in Ligeti's Continuum: A theoretical and

perceptual study' [Cambouropoulos and Tsougras, 2008], in which he further explores auditory streaming principles, this time in the context of modern composition. He attempts to show how the human mind organises sequences of notes into auditory streams.

Chew and Wu's 'contig mapping' method [Chew and Wu, 2005] is inspired by the perceptual principles of voice-leading defined by David Huron [Huron, 2001] and thus far gives the highest level of accuracy of current voice separation algorithms. The algorithm is based on the principle of pitch proximity, supported by perceptual research that suggests that humans are more likely to hear notes that are close in pitch as constituting a single auditory stream, and the concept that humans perceive the divergence of auditory streams rather than voice crossing [Deutsch, 1975]. The primary rules of the method are summarised as follows:

- Pitch proximity: take the shortest route to make voice connections
- Stream crossing: do not allow voices to cross
- Ideally restrict the number of voices to three or fewer (based on Huron's principle of limited density).
- Permit only one note per voice

Chew and Wu define four entities intrinsic to their computational implementation: a *note* is an object with pitch and duration properties. A *fragment* is a linear sequence of successive notes belonging to the same voice. A *contig* is a collection of overlapping fragments such that the number of voices present throughout the contig is constant. A *maximal voice contig* is a contig containing the maximum number of voices. The algorithm works by segmenting the data into a series of adjacent contigs, seeding the maximal voice contigs first by pitch order, and then, via a metrics system which awards penalty points for less preferred connections, propagates the musical voicing outwards via nearest neighbours. Voice connections are therefore made at contig boundaries; by definition, the number of voices in adjacent contigs is different. Chew and Wu deal with the problem of fluctuating quantities of simultaneous notes, which using

the one note per voice rule would mean a much larger number of voices in total for a single musical work, by omitting large chords from their computation. They justify their approach by asserting that large chords serve as 'statements of finality', and that these verticals 'masquerade as maximal voice contigs' [Chew and Wu, 2005]. The omission is questionable however, although omitting large chords avoids generating many more voices than Huron's preferred maximum, the question as to exactly how such large chords should be voiced remains unanswered. Moreover, there are many instances in contrapuntal corpuses where large chords appear within the context of the piece and not simply at the end, suggesting that these chords are an integral part of the music, and by implication therefore, the voicing of the music.

Chew and Wu report an average voice consistency, (a measure of the average proportion of notes to have been assigned by the algorithm to the same voice), of 88.98% when applied to polyphonic keyboard music by J. S. Bach. Recent work based on a replica of Chew and Wu's method by Ishigaki improves the average voice consistency measure to an accuracy of 92.21% when the process of connecting contig boundaries is prioritised to prefer making connections at boundaries where the number of voices is increasing rather than decreasing [Ishigaki et al., 2011].

3.3 Chord Recognition

Automatically classifying keys and chords using a computer program to process digital data faces several common problems. The majority of systems process digital data formats such as MIDI, which generally does not supply the enharmonic pitch spelling and voicing information that would be levied by a human analyst performing the same task. For both key and chord extraction, a key issue is that of ambiguity, which can be defined as musical situations that present too little information to draw a conclusion with any degree of certainty, or situations presenting too much information and thus presenting a range of possibilities. Rohrmeier [2007] gives an example of the latter in relation to key finding, citing the chord sequence C-G-C-G, which could equally be interpreted as being in the key

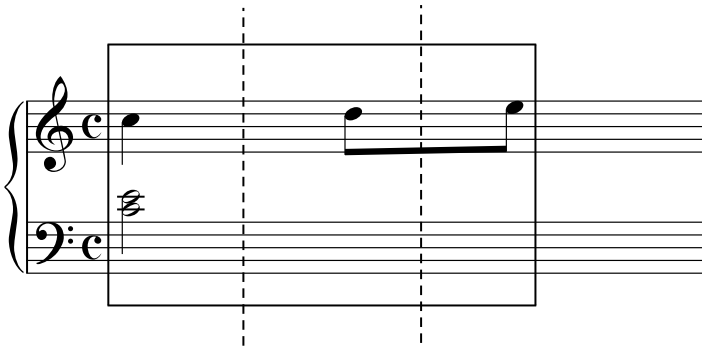


Figure 3.1: Reproduction of Pardo and Birmingham Excerpt.

of C major or G major. With chord extraction, the same issue of ambiguity presents itself in relation to pitch content. Mechanisms need to be put in place to ascertain which pitches in a presented group are chord tones, and which ones are not. Chord extraction from pitch class sets is a difficult problem, either involving the creation of complex interlocking rules, or some other method of pitch to chord label inference. Many researchers simplify the problem by processing only pitches which occur on the beat, ignoring pitches with onsets that are on the offbeat (e.g. [Ponsford et al., 1999]), however as discussed in section 6.1 it is erroneous to assume that pitches occurring on the beat are always chord tones.

Pardo and Birmingham [2002] describe a pitch counting algorithm to automatically recognise chords from symbolic data. Importantly this part of the method is not relying on musical rules but derives information from pitch weight evidence. The method is based on a pre-defined dictionary of pitch class set style chord templates, representing triads and two types of sevenths. The dictionary is populated by creating a pitch class set for each type of defined chord template on pitch class from 0 (C), ascending chromatically, to 11 (B). The templates are fully factored, i.e. every note member of a chord is represented, for example, the major triad on C, {C, E, G}, is defined as pitch class set {0, 4, 7}, and the seventh on G, {G, B, D, F}, is defined as {7, 11, 2, 5}. The chord algorithm determines the weight of individuals notes in a given segment by counting the number of fractional segments the note is present within. A fractional segment

is a sub-segment of the main segment denoted by a new note onset. For example, Figure 3.1 reproduces the example segment used in [Pardo and Birmingham, 2002], containing five notes, two minims, C and E in the bass staff, and a treble voice progression C (crotchet), D, and E (quavers). There are three fractional, or sub-segments, within the main segment. These are shown by dotted lines in the figure. A count of pitch presence in the fractional segments results in note weights of 3 (minim C), 3 (minim E), 1 (crotchet C), 1 (quaver D) and 1 (quaver E) respectively. The method therefore does not correlate to absolute duration but is a relative measure of durational strength. The note weights are then compared to each and every pitch class template in the dictionary, and a count is made of positive evidence, negative evidence and misses, as described in the list following:

- *positive evidence* = the sum of the weights of pitch classes matching a template element
- *negative evidence* = the sum of the weights of the pitch classes not matching a template element
- *misses* = sum of the template elements not matching the pitch classes.

A score for the pitch classes when compared to the template is calculated by deducting the sum of negative evidence and misses from the positive evidence values.

Three preference rules are used to choose between multiple pitch class templates producing an identical score:

- prefer templates whose root pitch class has the greatest weight of notes present in the template
- prefer templates having a higher prior probability
- in the case of diminished 7ths prefer the template whose bass note moves down by a semitone in the following chord.

An issue facing all computational work in this field is access to ground truth data against which to measure results, consequently a significant

contribution made by Pardo and Birmingham [2002] is their objective evaluation of results against the Kostka and Payne corpus [Kostka and Payne, 1984]. The corpus contains a range of excerpts of music including works by Bach, Beethoven, Haydn, Mozart, Schubert and Tchaikovsky. They achieve levels of accuracy ranging from 75.81% to 88.65% depending on the segmentation method used. In this work they also detail complex approaches to segmentation of the data, widely acknowledged to be a very hard problem to solve computationally. Segmentation can be defined as the temporal division of music data, such that each segment corresponds to a single chord. The issue is complex because harmonic rhythm in music (i.e. rate of chord change), fluctuates considerably throughout a single work and is also subject to differing opinion amongst annotators. Annotations of chords and keys by human musicians are often generalised to higher level segments so that fractional level chords are not labelled. Harmonic rhythm is discussed in more detail in subsequent chapters, see for example section 5.5.1. A great deal of harmony researchers avoid the issue of segmentation altogether by basing experiments on the homophonic Bach chorales ([Raphael and Stoddard, 1984, Rohrmeier, 2007]) which predominantly follow a harmonic rhythm of a crotchet beat.

Maxwell [1992] describes the implementation of a LISP based expert system to perform harmonic analysis. Maxwell states that the two main problems when constructing rules for harmonic analysis are knowing how to choose which vertically coinciding groups of notes are chords, and secondly where to demarcate the boundaries between chord groups. The system defines a complex set of 55, interacting, specific rules by which to recognise consonant and dissonant chord types including metrical accentuation and tertian stacking of notes. The implementation of the rule set is tested on J. S. Bach's *Six French Suites*, and produces figured bass labels for the test set. Maxwell discusses both the generally plausibility of the output, but also the problems of the results which range from a bias towards simpler triadic chord types, with some results missing clear examples of seventh chords, and the computational difficulty of achieving accuracy when there is abundant linear movement. For example, this may be when the movement of quavers which should be identified as passing

notes creates successive offbeat consonances. Maxwell's results are not measured against any ground truth; similarly to the work of [Raphael and Stoddard, 1984] the results are discussed using musical excerpts. However the lack of objective evaluation in this type of work makes it difficult to understand precisely how successful and accurate the work is. Maxwell highlights that the goal of this work was to encode rules intuitively, and to parallel the kind of mental processes used by a human analyst in the same task, and as such the work is a useful point of reference for further experimentation in the field of automatic chord recognition.

Research using harmony information to classify style and genre has been made possible by manually annotated data sets such as Chris Harte's Beatles chord annotations ([Mauch et al., 2007]). Manually annotating chord labels to this degree of accuracy is very laborious thus there are few complete sets of data to work with. Amelie Anglade has researched the automatic classification of genres of symbolic and audio music using harmony rules derived from manually annotated chord progressions ([Anglade and Dixon, 2008, Anglade et al., 2009, 2010]). Her research found that many chord progressions are generic across different genres and that the difference between genres based on chord data is subtle. The results demonstrate the necessity to capture intricacies of harmony usage beyond the commonality of basic triads and simple sevenths for style and genre recognition in music. Improving the level of richness of information about vertical sonorities, for example, to include the presence and compositional treatment of complex dissonance, as well as the use of chord progressions in combination with other features such as key or harmonic vectors, as described by Phillip Cathé may be more revealing [Cathé, 2010].

A great deal of research has been performed to classify chords directly from audio, for example [Mauch and Dixon, 2010, Harte et al., 2005, Fujishima, 1999]. Mauch concentrates on abstracting chords from audio representations of pop music using a dynamic Bayesian network that combines information about meter, key, chord and bass and treble chroma. Mauch reports that the use of contextual information improves chord recognition accuracy.

3.4 Statistical and Probabilistic Work

> in predicting a future stimulus, our best prediction would be the stimulus that has occurred most frequently in the past [Huron, 2007]

In his book 'Sweet Anticipation', David Huron moots the idea that statistical studies of music could yield novel insights into music. Chapter 5 of the book, 'The Statistical Properties of Music', is dedicated to this idea, and he demonstrates the applicability of statistics to music by showing that conclusions can be made about music from statistical studies. For example, the frequency of occurrence of melodic intervals for samples of music from ten cultures spanning Africa, Asia, Europe and America show that on a linear level small intervals predominate. Huron reiterates the idea mentioned in [Krumhansl, 1990] that listeners are also sensitive to the statistical regularities and distributions of tones in music, and that as a consequence there is a great deal of scope for further statistical, and by implication probabilistic, modelling of music. Hillewaere et al. [2009] have explored different statistical methods to differentiate style, either of composers or for folk song classification. They describe two methods of statistical analysis and compares the results of both when applied to the same data set. The first method, the 'global feature' approach, summarises a melody as a single feature vector; the second method is that of an 'event' model, which uses a sliding window to calculate the probability of a melody. The polyphony of the scores is not fully accounted for; the melodic lines are processed separately as if they were individual monophonic lines and the musical interaction of the voices is disregarded. Individuating Mozart from Haydn is clearly a very challenging task and it is suggested that work to elicit much deeper and complex musical features is necessary to successfully perform this task. Melodic intervals are easy to measure and work with, however basing experiments solely on melodic intervals thwarts the possibility of yielding deeper insights about a musical corpus, as shown in the variation between different modelling methods; for the folk song corpus the event model out-performs the global features, but in a similar experiment to differentiate between Mozart and Haydn String

Quartets [Hillewaere et al., 2010], the results are less determined. Conklin and Bergeron [2010] describes a method to combine features in order to discover abstract relations between contrapuntal parts. Their method of capturing melodic interactions between pairs of voices is applied to Bach chorale harmonisations proffers some success at eliciting distinctive patterns in the counterpoint.

David Temperley also strongly supports the idea of probabilistic studies, and has published a book about his work and that of others in this area [Temperley, 2007]. In the book Temperley describes Bayesian theory, reviewing current research and introducing a set of his own Bayesian algorithms which generate probability-based metre and pitch models, using the Essen folksong collection as his data set. Temperley's key finding model given a set of pitches performs reasonably well on annotated data, with results in the region of 80%. He points out that his probability models tend to perform less well than the rule based models used in the Melisma analyser. The models are completely statistical with no musicological rules applied. An important omission is that rhythm, bar and beat position information is not taken into account, and no account is made for structurally important notes. In Chapter 8 of his book, Temperley discusses the possibilities of applying Bayesian musicological modelling to transcription systems, saying 'it provides a natural way of bringing to bear higher-level musical knowledge'. He specifically cites the musical principle of pitch proximity as holding promise in this field of research. Partly related to the idea of pitch proximity (indeed a forerunner) is the formal musical principle of voice-leading.

Another area of musicological research which has been used to model music with some success is Markov modelling theory. Markov modeling has been used in a variety of contexts, such as tonality estimation and composer identification [Noland, 2009, Liu, 2002, Ryynänen, 2008].

Chapter 4

Key Estimation from Perceptual and Theoretic Data using Hidden Markov Models

Musicologists cite the harmonic language of a composer as a critical indicator of musical style period, as discussed earlier in this thesis (section 2.2). This includes: the range of chords used; types of chord progressions; the exploitation of consonance and dissonance; the frequency, style and methods of modulation; and key and key relationships. Translating such subtle, complex and inter-related musical phenomena into the rigorous terms required for computer processing is not a straightforward task, not least because of the interpretive nature of musical harmony. Music analysts will vary quite markedly in their views on the structural importance of notes, chord definition and chord function, key area, and tonality. In computational work, the exact moment of key change has been shown to be difficult to pinpoint, [Rohrmeier, 2007, 2011], due to a phenomenon he terms 'revision', whereby chords of 'dual function' belonging to both the previous key and the new key are reinterpreted when the new key becomes apparent.

This chapter describes a set of experiments to automatically detect key and modulation in J.S. Bach chorales from audio and MIDI data formats. The process comprises a number of stages. The input music data is initially processed into a series of temporally segmented notegroups that are automatically classified to generate a discrete output sequence of chord designations. The chord sequences are used as the input to various hidden

Markov models (HMMs) [Rabiner, 1989] that are constructed to detect key and key change in the chord sequences. The first set of models are built with the aim of systematically realizing some of the fundamental components of Schönberg's harmonic theory [Schönberg, 1922]. The second set of models test data from Krumhansl's book chapter describing perceived relationships of chords and keys in tonal hierarchies [Krumhansl, 1990], analogous to previous work by [Noland, 2009]. The models based upon heuristically derived data representing music theory and the models based on the results of Krumhansl's perceptual experiments are compared and conclusions are drawn about the overall approach. A final stage is the formulation of functional harmony labels for the score by combining the output key sequence of the HMM with the chord input sequence.

To the author's knowledge, this is the first study which utilizes polyphonic music transcription for systematic musicology research. We consider that such collaborative work has exciting potential, both for the advancement of automatic transcription, and for computational musicology.

4.1 Chorale Corpus

Twelve J.S. Bach chorales were selected from `www.jsbchorales.net`, which provides organ-synthesized recordings along with aligned MIDI reference files. The use of Bach chorales in computational musicology is extremely common due to their homophonic nature, which reduces the need to consider voice-leading elements and allows simple chord matching processes to be successful. The list of the chorales employed for the key detection experiments can be seen in Table 4.1. Sample excerpts of original and transcribed chorales are available online[1].

[1] `http://www.eecs.qmul.ac.uk/~emmanouilb/chorales.html`

Table 4.1: The list of organ-synthesized chorales used for key detection experiments.

	BWV	**Title**
1	1.6	Wie schön leuchtet der Morgenstern
2	2.6	Ach Gott, vom Himmel sieh' darein
3	40.6	Schwing dich auf zu deinem Gott
4	57.8	Hast du denn, Liebster, dein Angesicht gänzlich verborgen
5	85.6	Ist Gott mein Schild und Helfersmann
6	140.7	Wachet auf, ruft uns die Stimme
7	253	Danket dem Herrn heut und allzeit
8	271	Herzlich tut mich verlangen
9	359	Werde munter, mein Gemüte
10	360	Werde munter, mein Gemüte
11	414	Danket dem Herrn, heut und allzeit
12	436	Wie schön leuchtet der Morgenstern

4.2 Music Transcription

For the transcription of audio, we used the signal processing based transcription of [Benetos and Dixon, 2011]. Since the application of the transcription system concerns chorale recordings, the pitch range was limited to C2-A♯6 and the maximum polyphony level was restricted to 4 voices. The pitch candidate set that maximizes the score function is selected as the pitch estimate for the current frame. Finally, note offset detection is also performed using HMMs trained on MIDI data from the RWC database [Goto et al., 2003]. The recordings are synthesized, therefore the tempo is constant and beats can be estimated directly from the onset detection functions as described in [Benetos and Dixon, 2011]. The pitches in the time frames between two beats are estimated by the frame level data, computing the pitch salience function, resulting in a series of chords per beat. Transcription accuracy is 33.1% using the measure of [Benetos and Dixon, 2011], which also takes into account note durations, hence the low value. An example of the transcription output of BWV 2.6 '*Ach Gott, vom Himmel sieh' darein*' is given in Figure 4.1.

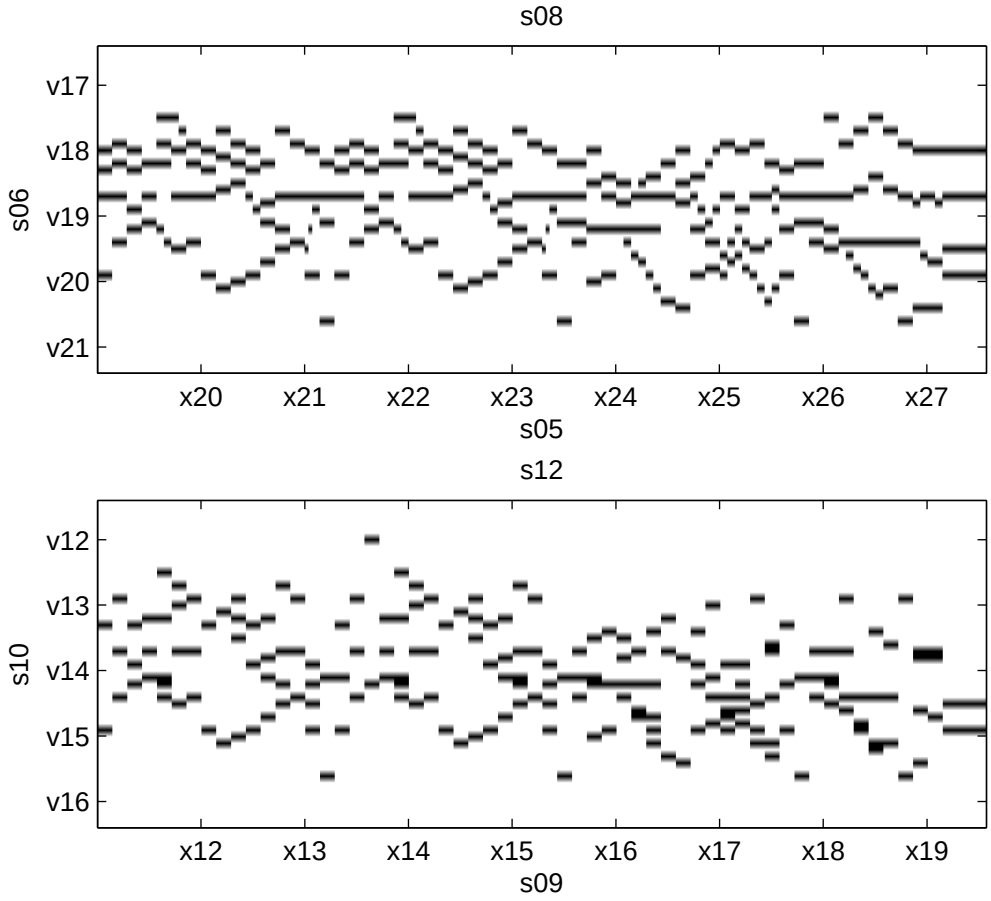


Figure 4.1: (a) The pitch ground-truth of BWV 2.6 '*Ach Gott, vom Himmel sieh' darein*'. (b) The transcription output of the same recording. The abscissa corresponds to 10 ms frames.

4.3 Chord Recognition

Transcribed audio, and for comparison, ground truth MIDI files, are segmented into a series of beat level vertical notegroups according to onset times. Notes which occur simultaneously or overlap in time are therefore grouped. The pitch values within a group are converted to pitch classes 0 to 11, (0=C, 1=C♯ etc), but the original pitch order (low to high) is preserved. It is also possible to have empty notegroups, representing rest values rather than notes. MIDI pitch numbers are kept in order of ascending value, with an assumption that the lowest note is the bass voice. All

instances of repeated pitch classes except the lowest (first) are removed to create a unique ordered set. For example, MIDI pitches {53, 57, 60, 65}, (bass, tenor, alto, soprano) would become pitch classes {5, 9, 0, 5} (modulo 12), which would become the unique ordered group {5, 9, 0}. The Bach chorales most commonly have a harmonic rhythm, (i.e. rate of harmonic change), of a crotchet beat, consequently for these experiments the vertical notegroups are organized into higher level groups which contain all of the notes present within this timing division. Thus, if the four notes of MIDI pitch {53,57,60,65} occurred on the first beat of the bar, (a metrical position of 1), but the MIDI note of pitch 65 (soprano voice) changed on the quaver offbeat, metrical position 1.5, (to MIDI pitch 63), the complete ordered group of pitch classes within the crotchet beat would be {5, 9, 0, 3}.

The notegroups are classified using a chord dictionary of templates of ordered sets of pitch classes, (e.g. a C major chord is {0, 4, 7}). The chord designations used for the chord dictionary are taken directly from Schönberg's 'Theory of Harmony' [Schönberg, 1922]. Schönberg defines unequivocally the diatonic triads, sevenths, ninths, elevenths, and 'vagrant' chords per major or minor mode. In the minor mode, the sixth and seventh degrees may be either raised or lowered by a semitone, more usually raised in ascending melodic/motivic lines leading to the tonic, and flattened on the descent. Schönberg makes a clear distinction between the major and the minor modes, which are dealt with separately throughout the text, and he is specific about the chords which identify them. The presence of the raised or flattened sixth and seventh degree in the minor mode results in more possible chord configurations than for the major mode. The only triad in the minor mode which is unaffected by this variable use of the sixth and seventh degree is the tonic triad. With regards to chords containing an added seventh, there are two possible configurations on each degree of the scale, and four possible configurations on the seventh scale degree. To see the diatonic triads and sevenths defined by Schönberg please refer back to Section 2.2.1, specifically Figures 2.1, 2.2, 2.3, and 2.4.

The program generating a chord dictionary commences with a pitch

class representation of the scale types for the full set of major and minor scales. For example, C major is [0, 2, 4, 5, 7, 9, 11]. The minor scales contain optional values for the sixth and seventh degrees which are bracketed, for example C minor is [0, 2, 3, 5, 7, [8, 9], [10, 11]]. The types are transposed across the full set of semitones from 0 to 11. The chord content of the dictionary is built using a set of chord interval profiles that are used to generate each chord type per scale degree. For example, the successive interval profile of a major triad is {4, 3} semitones - a major third plus a minor third successively from the root. In order to accommodate the different HMM's related in the subsequent sections of this chapter, two different chord dictionaries are used. The first chord dictionary represents the full set of major, minor, diminished, and augmented triads for the 12 major and 12 harmonic minor scales, resulting in a chord dictionary of 48 chord types. The second chord dictionary is more comprehensive, and represents the full range of diatonic triads and sevenths as defined by Schönberg and shown previously in Section 2.2.1. This dictionary also defines the selection of chromatic triads defined in [Kitson, 1920] and shown in Figure 4.2 in order to provide a comprehensive range of common chords per key. The result is 21 chord types defined for the major key, and unique 30 chord types for the minor key. The chords are then transposed into the full set of 12 major and 12 minor keys creating a dictionary of 132 chords. Ninths, elevenths and thirteenths are not represented in the dictionary due to the additional complexity of chord matching and chord template representation for extended chords. For example, a thirteenth on the root note of G would contain all tones from the root to thirteenth: {7, 11, 2, 5, 9, 0, 4}, or {G, B, D, F, A, C, E}. A thirteenth rarely appears in this form and typically omits some chord tones, such as the 9th, 11th and 5th. This creates ambiguity when matching and labelling such chords and increases the likelihood of error. In addition, the role of such chords in the definition of key is considered to be weaker than that of the primary triads and sevenths in the context of common practice harmony (for example see [Kramer, 1981] on nineteenth century harmony).

To obtain a discrete set of chord symbols per transcribed audio and MIDI file, every unique pitch class contained in a beat segment is passed

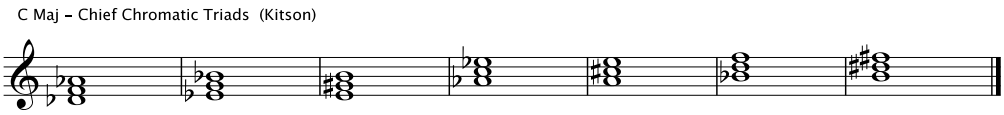


Figure 4.2: Kitson Chromatic Triads [Kitson, 1920].

through to the chord template matching method and is matched to the contents of one of the chord dictionaries. For the corpus described two chord sequences per chorale are produced, one which is limited to triads only, and another which is classified using the larger chord dictionary containing sevenths and chromatic chords. During the classification process, no durational or rhythmic weights, chord tone doubling preferences, or preference rules relating to passing or neighbour notes are used. The complete unique set of pitch classes present throughout the segment are tested. This initially simplistic approach is deliberate, primarily because the derivation of rules which would apply generically across many different polyphonic works is a complex area requiring further, and detailed research. In addition, all of the notes, including those occurring at fractional positions within the bar, may still be very much a part of the harmony. Bach frequently introduces the seventh note of the V7 chord on the offbeat, a gesture which would be considered to be a simple elaboration of a V7 chord and not a chord V, by many musicologists Kitson [1907]. By including the offbeat notes in the group, the chord designation is correctly assigned to a V7 and not a chord V as would have been the result otherwise. The inclusion of all tones per segmentation value restricts the scope of the method to mostly homophonic music: presenting all of the notes per segment in complex polyphonic keyboard music would result in many 'no-chord' matches without the introduction of the weighting of notes in the algorithm (see below for further discussion).

The chord matching process undergoes a series of iterations to find the dictionary template or templates that most closely match the presented notegroup in terms of edit distance. An exact ordered match between the pitch classes of the segment and a chord template in the presented dictionary, (edit distance 0), would be, for example, a root position triad (e.g.

{0, 4, 7}). An unordered exact match, (edit distance 0.5), would be an inverted chord (e.g. {4, 7, 0}). The process continues, adding 1 for each insertion or deletion, up to a maximum edit distance of 2. Restricting the edit distance to 1 significantly decreases the number of notegroup to template matches and can result in too many 'no chord' values. Equally, allowing an edit distance of 2 or 3, results in the generation of increasing numbers of multiple possible template matches per notegroup. The maximum edit distance is therefore restricted to 2. In the event that there is a change in the harmonic rhythm, such as an increase of chord change frequency in the approach to a final or structurally important cadence, the matching process delivers inaccurate results due to the number of contradictory chord tones in the group for the larger segment. Therefore if a match is still not found for the notegroup for a segment, the offbeat notes are removed from the presented group and the match process is repeated with the set of notes which occurred on the beat. Removing the additional offbeat chords results in a matched sequence. If no match is found, the chord is returned as no chord.

Notegroups containing non-chord tones can be ambiguous as to chord designation, and in many instances there are multiple possible template matches per notegroup. For example, if only two notes matching the tonic and the fifth of a triad are present, the template will match both the major and minor triad. The HMMs require a discrete sequence of chord symbols, consequently groups of tones returning more than one possible chord classification are reduced to a single chord choice by the application of preference rules. The first rule retains a chord which matches a root position profile and those which have a different inversion profile are discarded if there are root position chords present. The second rule selects chord options on the basis of local context matching, which searches near neighbours in the sequence first of all previously and then following in the series, (the search range value is a parameter of the algorithm), for identical chord labels. If an exact match is found between one of the multiple chords in question and a nearby chord, this chord is selected. Context matching has proved to be highly effective for the corpus used. In the example of an ambiguous dyad containing a tonic and a fifth, in

the preceding or immediately following beats, a definitive matching triad is invariably present, allowing for an accurate selection. Pardo and Birmingham [2002] came across the same problem. Their program labelled a chord containing a C and a G as either C minor or C major. In the case quoted, a definitive label would have been found by local context matching. In the event that there are still multiple chord options remaining, the final rule applied is taken from [Pardo and Birmingham, 2002], and selects chords on the basis of which chord type has the higher prior probability, which they list as being major, dominant seventh, minor, diminished 7th, half diminished seventh, and diminished triad respectively. For the corpus this series of rules effectively reduces the sequences to a single chord option per segment.

To measure the competence of the chord labelling process, the automatically generated chord sequences are compared to hand annotated sequences that have been annotated by the author of this thesis. The chorales in the set have been annotated with ground truth chord labels which include sevenths, thus the accuracy of the triads-only sequences is not measured. Each pair of chord values in the two sequences is compared (hand annotated and automatically generated), and a difference measure is calculated by counting the number of exact matches. The final counts are normalised, resulting in a proportional measure of matched or mismatched values between the two files. If two chords differ, the Levenshtein distance is calculated for the two pitch class sets represented as strings, to find out the degree of difference between the automatically classified chord and the hand annotated chord. Many of the chord mismatches found are in fact extremely close pitch class set matches, for example, {t, 2, 5} compared to {t, 2, 5, 9}, (where t=10, and e=11 due to the requirement for each symbol to be represented by a single character), generating a Levenshtein difference of 1. The accuracy results and the average Levenshtein distance for the mismatches in the file are shown in Table 4.2.

A greater quantity of label mismatches are found with the transcribed files than with the symbolic MIDI files, due to the pitch and timing errors resulting from the transcription process. Total chord mismatches between the transcribed data and the hand annotated data (i.e. where there are

Table 4.2: Chord match results for transcribed audio and MIDI against hand annotated chords.

	MIDI		**Transcribed Audio**	
BWV	%Match	Levenshtein Avg	%Match	Levenshtein Avg
1.6	95.0	2.5	71.2	2.0
2.6	90.0	1.7	70.0	2.1
40.6	84.4	1.5	57.8	2.0
57.8	85.2	2.4	64.8	2.0
85.6	85.7	1.0	48.2	2.1
140.7	96.6	1.0	72.1	2.3
253	82.5	2.0	65.0	1.9
271	83.1	1.0	67.7	1.9
359	78.1	1.2	76.6	2.5
360	84.4	1.1	71.9	2.4
414	73.3	0.9	68.3	2.6
436	93.8	1.4	67.5	1.8
Avg:	86.0	1.5	66.7	2.1

no pitches in common between the two pitch class sets), indicate an error in timing or quantisation. The greatest difficulty posed to the chord algorithm by the transcribed data is the frequent presence of dyads rather than triads in the groups. Resolving a dyad correctly is not straightforward; if the dyad is a third apart, this could imply either the upper or lower portion of a triad; equally, a dyad a fifth apart could be either a major or a minor triad. The transcription algorithm has a low false alarm error rate and a high mis-detection rate, consequently the transcription process produces output that assists the chord method where the MIDI data poses problems; groups with many non-chord tones, or notegroups containing complex chord tones unrepresented in the chord dictionary, are captured from the transcribed data as simple triads whereas the MIDI data may result in a 'no chord' value or erroneous label. Complex chords are less adaptable to the pitch class set match approach due to the fact that internal tones must be omitted from such chords to fit with four part harmony. The majority of errors in the MIDI data result from suspended and passing notes. Consequently the chord sequences obtained from transcribed audio do not contain any 'no chord' values either for complex chord

sequences or triadic sequences. Out of the complete set of 12 files, four files contain 'no chord' values in the resulting sequence, with chorale 57.8 producing the highest number of unrecognised chords due to the number of suspensions. The chord sequences containing complex chord symbols feature a slightly increased proportion of 'no chord' values (2.58% across the complete set), compared to the triadic sequences (1.1%).Overall, the accuracy levels, shown in Table 4.2, when compared to the ground truth files are in the upper range of the results reported in [Pardo and Birmingham, 2002]. (The transcribed audio achieves an average of 66% correct of the hand annotated data.)

4.4 Key Modulation Detection

The method chosen to deduce key from the chord sequences is a hidden Markov model (HMM). An HMM is a probabilistic model in which a series of hidden states are inferred from an observable sequence of data by calculating Bayesian probability values [Rabiner, 1989]. In the model the observation sequence $O = \{o[n]\}$, $n = 1, \ldots, N$ is given by the output of the chord recognition algorithm in the previous section. The observation matrix ($\mathbf{B}$) therefore defines the likelihood of a key given a chord. Likewise, the hidden state sequence which represents keys is given by $S = \{s[n]\}$, where $s[n] \in \{1, 2, ...24\}$. Each HMM has a key transition matrix $\mathbf{A} = P(s[n]|s[n-1])$ (representing the 12 major and 12 minor keys, as shown in Table 4.3), which defines the probability of making a transition from one key to another. The keys are ordered in accordance with the two lines of 5ths so that a move from one key to a close neighbour on the circle is apparent from the key numbering.

For a given chord sequence, the most likely key sequence is given by:

$$\hat{S} = \arg\max_{s[n]} \prod_{n} P(s[n]|s[n-1])P(o[n]|s[n]) \tag{4.1}$$

which can be estimated using the Viterbi algorithm [Rabiner, 1989]. In Figure 4.3, the graphical structure of the employed HMM model is shown.

Table 4.3: Representation of Keys.

s[n]	Key
1	C Maj
2	G Maj
3	D Maj
4	A Maj
5	E Maj
6	B Maj
7	F♯ Maj
8	C♯ Maj
9	A♭ Maj
10	E♭ Maj
11	B♭ Maj
12	F Maj
13	A Min
14	E Min
15	B Min
16	F♯ Min
17	C♯ Min
18	G♯ Min
19	D♯ Min
20	B♭ Min
21	F Min
22	C Min
23	G Min
24	D Min

HMMs are used for a wide range of modelling purposes with good reason. As highlighted by Marsden in his discussion regarding the perception of musical voices, 'if a model serves as a methodological device, it is crucial that is should be comprehensible and its workings be clear' [Marsden, 1992]. In addition to this, Marsden goes on to explain, a model should also be 'predictable', in that 'the representation of its knowledge is precise, and it should be 'extensible', i.e. it should be possible to add further rules or knowledge to the model with ease. The HMM framework allows for all of these: clarity, predictability, and extensibility.

An HMM has been used previously to infer the overall key of a piece using Krumhansl's perceptual data [Noland, 2009]. Krumhansl performed

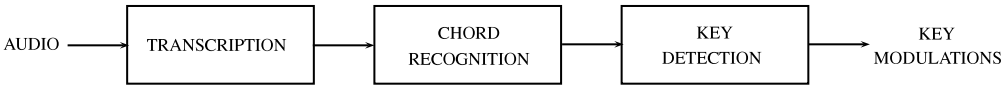


Figure 4.3: Graphical structure of the employed HMM for key modulation detection.

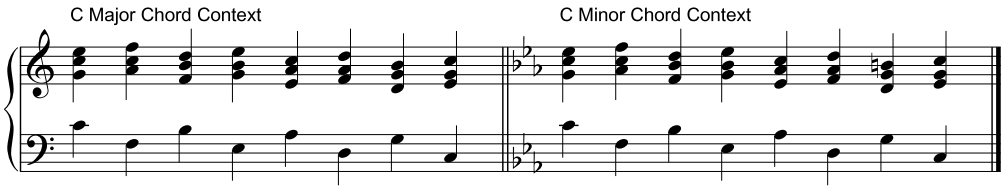


Figure 4.4: Major and minor chord contexts used in the Krumhansl harmonic hierarchy experiments [Krumhansl, 1990].

perceptual experiments with moderately trained listeners to test the structural significance of chords in tonal contexts [Krumhansl, 1990]). Two experiments were performed, the first with 10 listeners, and the second with 12. In the first experiment the key context given was the major or harmonic minor scale, in the second experiment, the key context was a major or minor chord sequence, reproduced in Figure 4.4. In both experiments, the key context was followed by a single chord, and the listeners were asked to rate 'how well the chord fit' the preceding context, using a ratings scale from 1 - 'fits poorly' to 7 - 'fits well'. In the first experiment the context was followed by each of the major, minor, diminished and augmented triads. Due to minimal variation in the ratings for augmented chords in the this experiment, the augmented chords were omitted from the second experiment. The chord ratings data produced by the two experiments is shown in Table 4.4.

Despite the success of using this data previously, a curious feature of the data produced by the experiments is that the chord ratings do not ratify music theory in terms of clearly associating chord/key membership. (Consider that a major key consists of three major triads (I, IV, V), three minor triads (II, III, VI) and one diminished triad (VII).) In the C major

Table 4.4: Chord ratings resulting from harmonic-hierarchy experiments [Krumhansl, 1990].

	Key Context			**Key Context**			**Key Context**	
Chord ↓	**C Major**	**C Minor**	Chord ↓	**C Major**	**C Minor**	Chord ↓	**C Major**	**C Minor**
C Maj	6.6 (I)	5.30	C min	3.75	5.90 (i)	C dim	3.27	3.93
C♯/D♭ Maj	4.71	4.11	C♯/D♭ min	2.59	3.08	C♯/D♭ dim	2.70	2.84
D Maj	4.60	3.83	D min	3.12 (ii)	3.25	D dim	2.59	3.43 (ii)
D♯/E♭ Maj	4.31	4.14 (III)	D♯/E♭ min	2.18	3.50	D♯/E♭ dim	2.79	3.42
E Maj	4.64	3.99	E min	2.76	3.33	E dim	2.64	3.51
F Maj	5.59 (IV)	4.41	F min	3.19	4.60 (iv)	F dim	2.54	3.41
F♯/G♭ Maj	4.36	3.92	F♯/G♭ min	2.13	2.98	F♯/G♭ dim	3.25	3.91
G Maj	5.33 (V)	4.38 (V)	G min	2.68	3.48	G dim	2.58	3.16
G♯/A♭ Maj	5.01	4.45 (VI)	G♯/A♭ min	2.61	3.53	G♯/A♭ dim	2.36	3.17
A Maj	4.64	3.69	A min	3.62 (vi)	3.78	A dim	3.35	4.10
B♭ Maj	4.73	4.22	B♭ min	2.56	3.13	B♭ dim	2.38	3.10
B Maj	4.67	3.85	B min	2.76	3.14	B dim	2.64 (vii)	3.18 (vii)

context of Krumhansl's experiments, all of the twelve major triads, irrespective of which note is the chord root, are rated as inferring the key of C major more highly than the diatonic chords actually belonging to the key of C major, (these may be minor or diminished in profile). Krumhansl refers to this as the 'chord type effect', stating that 'in the major key context, listeners strongly preferred major chords over minor and diminished chords', and that this perhaps may be accounted for by the relative degrees of consonance, with listeners preferring the most consonant chord types [Roberts and Shaw, 1984]. Krumhansl's experiments appear to indicate that perceptually, any major chord is more indicative of any major key, than the diatonic chords which make up that key, simply because it sounds major. An alternative interpretation, is that the data evidences *chord similarity* ratings rather than *key fittingness*, or the extent to which a chord is perceived as sounding as though it fits within a key [2]. It is possible that if the question asked of the listeners had been phrased differently, for example 'to what extent does the chord sound as though it is in the same key as the context?', a different set of ratings would result. The ratings data certainly appears to group chords by type more than it supports key membership.

From the perspective of music theory and common compositional practice, the data is therefore counterintuitive and one would anticipate inconsistent results when used with common practice musical works. It is hypothesized therefore, that populating the HMM matrices with data heuristically derived from music theory, will result in an improvement in key detection accuracy. Because Schönberg unequivocally defines the diatonic chords for the major and minor mode, a premise upon which the design of the theory models are based, is that the observation matrix should be able to strongly indicate key because the chord values are so closely derived from core harmony principles. To test this hypothesis, three HMMs are constructed based on Schönberg's harmonic theory, and two more are constructed embodying Krumhansl's data.

Krumhansl uses the interlocking pattern of quantified chord functions

[2]The author would like to acknowledge the examiner Alan Marsden for this idea.

Table 4.5: Correlations between harmonic hierarchies [Krumhansl, 1990]

Key	**C Major**	**C Minor**
C Major	1.000	.738
C♯/D♭ Major	-.301	-.224
D Major	-.141	-.320
D♯/E♭ Major	-.013	.405
E Major	-.139	-.256
F Major	.297	.194
F♯/G♭ Major	-.407	-.281
G Major	.297	.175
G♯/A♭ Major	-.139	.123
A Major	-.013	-.286
A♯/B♭ Major	-.141	.013
B Major	-.301	-.298
C Minor	.738	1.000
C♯/D♭ Minor	-.298	-.373
D Minor	.031	-.189
D♯/E♭ Minor	-.286	.072
E Minor	.123	-.096
F Minor	.175	.245
F♯/G♭ Minor	-.281	-.321
G Minor	.194	.245
G♯/A♭ Major	-.256	-.096
A Minor	.405	.072
A♯/B♭ Minor	-.320	-.189
B Minor	-.224	-.373

to derive a map of key distances from the chord ratings, shown in Table 4.5. The key distance values produce a set of values which form a representation of the circle of fifths, (please see section 2.2.5).

4.4.1 Model Definitions

Five observation matrices (**B**) and four key transition (**A**) matrices are constructed in total. Three of the observation matrices are derived from music theory, and are designed to represent and test Schönberg's theory with regard to the chord membership of the 24 major and minor modes [Schönberg, 1922] (see section 2.2.1). Two further observation matrices use data from Krumhansl's perceptual experiments [Krumhansl, 1990]

Table 4.6: Rules for Schönberg observation matrix.

Feature	**BSchP**	**BSchCh**
Diatonic chord	1	2
Scale degree	2	1
Dim/aug scale degree	1	0.5
Ambiguous scale degree	1	0.5
Dim/aug ambiguous scale degree	0.5	0.25
Tonic chord	1	1

transposed across all 24 keys. The four different key transition matrices, (defined below), are used in conjunction with all five of the observation matrices.

Observation Matrices from Music Theory

The extent to which a chord infers a key is modelled heuristically in the music theory observation matrices. The intention is to logically produce a set of musically plausible chord rankings per key across the full range of chords observed. The diatonic chords of a key are all indicative of the home key; progressions containing chords II or IV with V or V7 are strongly indicative of the home key because they would have to be chromatically altered to imply a different key [Piston, 1983]. Similarly, the tonic triad, although it could be a member of several keys, tends to be prominent in the establishment of a tonal centre. Such chords may therefore be expected to rank highly compared to the lower values achieved by less characteristic chords. The relationship and interdependencies of individual tones, chords, and keys to human cognitive processing of tonality is not well understood. Consequently, to arrive at a score for a chord in relation to a key, points are given for both tone and chord properties. These include, points for each constituent tone per scale degree membership, partial points for ambiguous scale degree membership (i.e. 6th and 7th degrees in the minor key), for tonic chord status, and for being defined as a diatonic chord for the key by Schönberg. The points are then summed to give a total score for the chord in that key context.

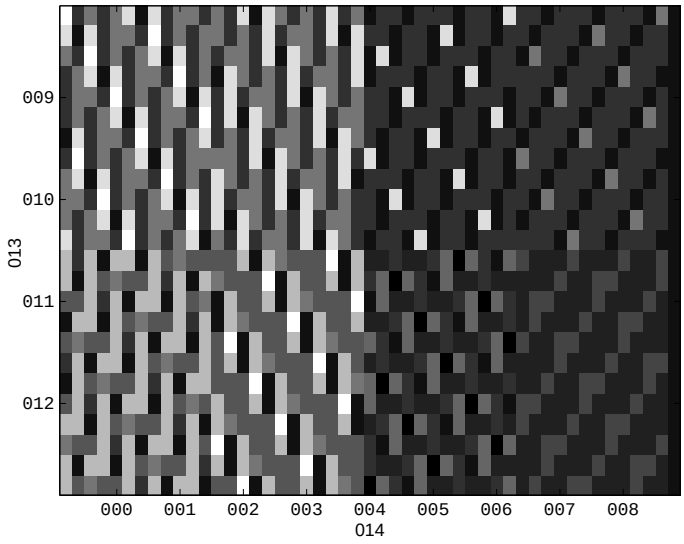


Figure 4.5: The *BSchCh* observation matrix

Two of the Schönberg observation matrices symbolise the complete set of major, minor, diminished and augmented triads plus a 'no chord' value, resulting in a total of 49 possible chord symbols. The two matrices are weighted differently, in order to test Parncutt's psychoacoustical work suggesting that chords are heard as having singular identities which are prior to the constituent pitches [Parncutt, 1989]. Matrix *BSchCh* therefore assigns double points to the diatonic chord as whole and gives single points for individual tones, whereas *BSchP*, gives double points to constituent tones and single points for diatonic chord status. The precise rules and values used are listed in Table 4.6 and an image of the normalised *BSchCh* matrix can be seen in Figure 4.5.

For example, the chord rating for a C major triad in the key of C major for *BSchP* would be as follows:

- C,E,G, tonic chord = +1
- C,E,G, three diatonic scale degrees = 2+2+2
- C,E,G is listed by Schönberg as one of the diatonic chords = +1

- Chord total = 8.

The third observation matrix *BSch7* symbolises the full set of triads and seventh chords elucidated by Schönberg [Schönberg, 1922] resulting in 22 chord definitions for the major key, and 30 chords for the minor key. The disparity in chord quantity is due to the optional raising of the 6th and 7th degree in the minor mode. A total number of 132 unique pitch class sets plus a 'no chord' value are defined, bringing the total number of possible chord observations to 133.

The values assigned to each chord in the *BSch7* model are the same as those used for *BSchP*. In this model, the value for the dominant seventh of C major would be:

- G,B,D,F, four diatonic scale degrees = 2+2+2+2
- G,B,D,F, is listed by Schönberg as one of the diatonic sevenths for C major = +1
- Chord total = 9.

The dominant seventh chord is the highest signifier in the matrix for its key, satisfactorily articulating common practice in tonal harmony.

Observation Matrices from Music Perception

The perceptual observation matrices symbolise the same chord set as the previously described triad based Schönberg models. The four triad based models therefore process identical chord sequences, allowing a direct comparison of the models based on music theory against those based on perceptual data.

The first matrix *BKrumOrig* is formulated using Krumhansl's chord ratings (Table 4.4, similar to previous work [Noland, 2009] with the difference that all of Krumhansl's chord data is used). In the absence of data for augmented triads, these plus the 'no chord' value are given a uniform low value of 1.0. As an experiment, a second observation matrix *BKrumMod* is also created, in which the apparently contradictory values for minor chords in the major key context which are part of the key, are swapped

with the major chord values which are not part of the key. For example, in the C major context, the values for the D major chord are swapped with the value for the D minor (chord II), E major with E minor (chord III), A major with A minor (chord VI), and B major with B diminished (chord VII). Performing this swap leads to disproportionately high values for the remaining major chords which also belie the home key without a parallel minor or diminished chord with which to exchange the rating. Such chords have 1 subtracted from their rating value to bring the data more in line with the swapped changes, for example the chord rating of 4.36 for F♯ major becomes 3.36. The values for minor chords in the minor key context in this model are left unmodified.

Key Transition Matrices

Four different versions of the key transition matrix are formalised and used in conjunction with all five of the observation matrices. The first matrix *ANeutral* is neutral, so that a move to any key is equally likely. The second transition matrix *AKrum* features Krumhansl's correlations between key profiles [Krumhansl, 1990] summed with 1. The third and fourth matrices, referred to as *ASchEq*, and *ASchNL* respectively, are implementations of Schönberg's table of key circles, in which seven circles of increasing key distance from a given tonic are delineated [Schönberg, 1922]. Using pitch class set representations there are six unique circles only, the seventh containing the enharmonically equivalent keys of previous circles. Therefore the *ASchEq* subtracts an equal value of 0.25 for each key circle, commencing with an upper boundary of 2.0, and moving through the relative minor and then each successive circle, ending on the 6th circle. The *ASchNL* implementation uses an exponentially decreasing value, halving the deducted value for each circle. In *ASchNL* therefore, the numeric distance between the first circle and the sixth circle is smaller than the distance between the same two circles in the *ASchEq* matrix. For all key transition matrices except the neutral matrix, the central diagonal is weighted by adding the value of 1 to give a small preference to stay in the current key. This is to model the human expectation that a chord

sequence is most likely to continue in the current key unless there is clear evidence of modulation. Without this weighting the models changed key very frequently. Conversely, adding too much weight to the central diagonal e.g. a value of 5, influenced the models to remain in the key irrespective of strong chordal modulatory implications. The values were determined by repeatedly running the models and manually comparing outputs to the ground truth data. An example of the values for each transition matrix for the key of C major is shown in Table 4.7.

4.5 Evaluation

4.5.1 Metrics

To provide a rigorous measure of accuracy of the outputs of the HMMs, each key value in the output sequences is compared to the corresponding hand-annotated key, and values are calculated by which to measure the performance of the models. These values include an error rate (*Err*), a distance measure (*Dist*), a measure of modulation concurrency, which is the number of times the HMM sequence changes key at precisely the same moment as the hand annotated sequence, this value is expressed as a percentage in the results tables (e.g. Table 4.8), (*Conc*), and modulation percentage (*Mods*) are calculated. Given N_{diff} the number of differences between output key and hand annotated key, N_{len} the length of the sequence, N_{cmod} the number of concurrent modulations, N_{hmod} the number of hand annotated modulations, and N_{omod} the number of modulations in the output, *Err*, *Conc* and *Mods* are defined as:

$$Err = \frac{N_{diff}}{N_{len}}, \quad Conc = \frac{N_{cmod}}{N_{hmod}}, \quad Mods = \frac{N_{omod}}{N_{hmod}} \tag{4.2}$$

The distance value *Dist* captures both the number of differences and the extent of each difference relative to the circle of fifths when two key values are found to conflict. For example, the distance value between two keys on the same circle, i.e. its dominant, subdominant, or relative minor, is 1, whereas a key difference two fifths apart on the circle of fifths (in either direction) would result in a difference value of 2, and so

Table 4.7: Value sets for the four key transition matrices shown for the key of C major.

	C	**G**	**D**	**A**	**E**	**B**	**F♯**	**C♯**	**G♯**	**D♯**	**A♯/B♭**	**F**
AKrum	2.000	1.591	1.040	0.895	0.815	0.500	0.317	0.500	0.815	0.895	1.040	1.591
ASchEq	2.000	1.500	1.250	1.000	0.750	0.5	0.25	0.5	0.750	1.000	1.250	1.500
ASchNL	2.000	1.500	1.000	0.75	0.5	0.375	0.25	0.375	0.5	0.75	0.375	1.500
ANeutral	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	a	**e**	**b**	**f♯**	**c♯**	**g♯**	**d♯**	**a♯/b♭**	**f**	**c**	**g**	**d**
AKrum	1.651	1.536	0.842	0.631	0.702	0.492	0.346	0.598	1.215	1.511	1.241	1.237
ASchEq	1.750	1.500	1.250	1.000	0.750	0.5	0.25	0.5	0.750	1.000	1.250	1.500
ASchNL	1.750	1.500	1.000	0.75	0.5	0.375	0.25	0.375	0.5	0.75	1.000	1.500
ANeutral	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000

on. *Conc* is regardless of whether the actual key change matches or not. Finally, *Mods* shows the percentage of the number of modulations in the HMM sequences compared to the number of modulations in the hand annotated key sequences. This last value is considered to be important firstly because it gives an indication of the effectiveness of the models in tracking modulation, and secondly because the frequency of modulation in a musical work is an important indicator of musical style. The results tables show the mean of all of the normalised data.

4.5.2 Results of Triadic Models

The results for all combinations of key transition matrices and observation matrices for the triadic models are shown in Table 4.8.

Error rates range from 0.26 to 0.35 for the transcribed data and 0.20 to 0.33 for the MIDI data sets. When the results are ordered by error, key distance measure, or the number of modulations relative to the number of modulations in the hand annotated data, the Schönberg observation matrices expose a pattern of consistently higher accuracy levels than the perceptual data matrices. The key transition matrices, for both the music theory models and the Krumhansl model, are less easily distinguished. The *ANeutral* matrix gives the poorest performance overall.

Matching the exact moment of key change between the HMM and the hand annotated sequences is a predicament because the hand annotated sequences take into account phrasing; key designations of chords depend upon both previous and subsequent harmonic movement, i.e. at moments of key transition the chords belong to both the current and new key. This makes the exact moment of key transition ambiguous. In the hand annotated data, the precise changeover point is decided on the basis of non-harmonic phrasing information. The HMM has no phrase information encoded, hence it will change key solely on the basis of chord and key transition data. The models often display a key change timing lag approximately one beat behind the annotated data. The modulation concurrence results are therefore quite low overall, but they are significantly higher for the Schönberg observation matrices, with the combination of *AKrum*

Table 4.8: Key detection results for all combinations of observation (B) and transition (A) matrices for triad models: error average (Err), distance value for key differences average (Dist), percentage of modulation timing match (Conc), number of modulations as a percentage of hand annotated number of modulations (Mods). Ground truth MIDI and transcribed file sets.

B Matrix →	**BSchP**				**BSchCh**				**BKrumOrig**				**BKrumMod**			
A Matrix ↓	*Err*	*Dist*	*Conc*	*Mods*	*Err*	*Dist*	*Conc*	*Mods*	*Err*	*Dist*	*Conc*	*Mods*	*Err*	*Dist*	*Conc*	*Mods*
MIDI																
ANeutral	0.31	0.45	9.25	38.26	0.27	0.45	24.59	85.26	0.22	0.37	8.45	31.72	0.33	0.53	5.01	22.87
AKrum	0.23	0.33	33.01	87.25	**0.20**	0.34	**46.03**	120.84	0.28	0.40	15.66	87.24	0.26	0.35	13.31	59.34
ASchEq	0.21	0.32	32.81	85.66	0.21	0.31	43.05	109.18	0.27	0.35	15.66	109.18	0.25	0.33	16.72	52.68
ASchNL	0.21	**0.30**	29.38	72.47	**0.20**	**0.30**	38.54	113.79	0.26	0.36	15.66	83.70	0.28	0.36	17.06	55.52
Transcribed																
ANeutral	0.35	0.74	11.55	51.09	0.27	0.45	25.23	109.42	0.28	0.42	4.76	18.89	0.32	0.51	2.68	9.68
AKrum	**0.26**	0.42	22.31	78.63	0.30	0.54	**37.58**	132.59	0.30	0.47	7.82	52.98	0.31	0.47	2.68	33.63
ASchEq	**0.26**	0.41	23.54	87.38	0.31	0.56	36.07	124.84	0.31	0.53	7.82	53.67	0.30	0.52	6.50	34.26
ASchNL	**0.26**	0.39	28.40	81.72	0.30	0.47	33.86	118.57	0.31	0.53	7.82	56.31	0.31	0.54	5.80	33.00

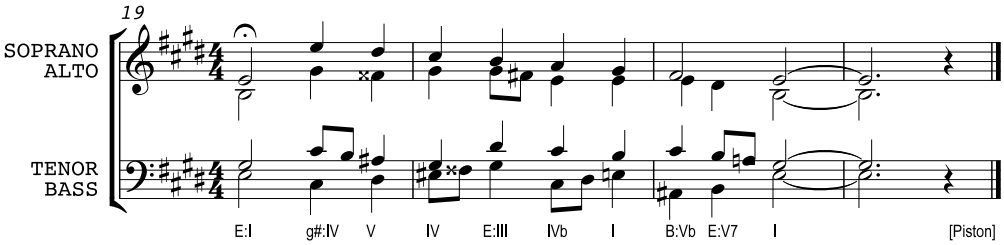
Table 4.9: Results of Student TTest comparing the level of deviation of error rates between transcribed data sets to MID data sets for each combination of observation (B) and transition (A) matrices for triad models.

B Matrix →	**BSchP**	**BSchCh**	**BKrumOrig**	**BKrumMod**
A Matrix ↓	*TTest*	*TTest*	*TTest*	*TTest*
ANeutral	0.85	0.40	0.43	0.82
AKrum	0.64	0.17	0.81	0.47
ASchEq	0.56	0.21	0.58	0.55
ASchNL	0.55	0.20	0.56	0.79

and *BSchCh* producing the most accurate moments of key change. Figure 4.6, which shows Piston harmony annotations alongside the results of the models [Piston, 1983], demonstrates the problem. The *BSchCh* observation matrix changes to G♯ minor on precisely the same chord as Piston and holds the key for four beats. *BSchP* also changes to the correct key, but a beat later. Although Piston annotates the G♯ minor triad of bar 20 in the excerpt as III of E major, it could equally be classed as chord I of G♯ minor, as per some of the HMM outputs.

The music theory data also appears to illustrate greater sensitivity to short digressions through other keys than the perceptual data. In terms of recognising global key, the perceptual models, which tend to stay in the home key when harmonic divergence is only for the length of a couple of beats, could be a preferred choice. If closer recognition of secondary dominants is desired, the music theory based models appear to be the more suitable option.

The key output accuracy using the transcribed audio for all models is encouragingly high when compared to the results for the MIDI data, achieving an average of 79% of the accuracy achieved for the ground truth data, despite the higher quantity of chord recognition errors for the transcribed audio. Table 4.9 shows a series of Student t-test results comparing the transcribed data sets to the MIDI data sets with a null hypothesis that the two sets have no statistical difference. All of the models produce a probability value that our null hypothesis is true with the *ANeutral BSchP* and *ANeutral KrumMod* giving the highest probability values to



ANeutral, BKrumOrig	E	E	E	E	E	E	E	E	E	E	E	E	E	E	E
ANeutral, BKrumMod	E	E	E	E	E	E	E	E	E	E	E	E	E	E	E
ANeutral, BSchP	g#	g#	g#	E	E	B	B	B	B	B	B	B	B	B	B
ANeutral, BSchCh	g#	g#	g#	E	E	B	B	B	B	B	B	B	B	B	B
AKrum, BKrumOrig	E	E	E	E	E	E	E	E	E	E	E	E	E	E	E
AKrum, BKrumMod	E	E	E	E	E	E	E	E	E	E	E	E	E	E	E
AKrum, BSchP	g#	g#	g#	E	E	B	B	B	B	B	B	B	B	B	B
AKrum, BSchCh	G#	G#	E	E	E	E	E	E	E	E	E	E	E	E	E
ASchEq, BKrumOrig	E	E	E	E	E	E	E	E	E	E	E	E	E	E	E
ASchEq, BKrumMod	c#	c#	E	E	E	E	E	E	E	E	E	E	E	E	E
ASchEq, BSchP	g#	g#	g#	A	A	A	E	E	E	E	E	E	E	E	E
ASchEq, BSchCh	g#	g#	g#	E	E	B	B	B	B	B	B	B	B	B	B
ASchNL, BKrumOrig	E	E	E	E	E	E	E	E	E	E	E	E	E	E	E
ASchNL, BKrumMod	c#	c#	E	E	E	E	E	E	E	E	E	E	E	E	E
ASchNL, BSchP	g#	g#	g#	B	B	B	B	B	B	B	B	B	B	B	B
ASchNL, BSchCh	g#	g#	g#	c#	B	B	B	B	B	B	B	B	B	B	B

Figure 4.6: Key outputs based on MIDI data of final bars of BWV 436, 'Wie schön leuchtet der Morgenstern', for all triad model combinations compared with Piston harmony annotations [Piston, 1983]

support this. The implication is that the transcribed audio is of sufficient quality for some musicological tasks based on predominantly homophonic textures. The transcription error rate for more complex contrapuntal textures would need to be improved.

4.5.3 Results of Sevenths Model

The results for the *BSch7* model in combination with all four key transition matrices are shown in Table 4.10. This more complex HMM containing 133 chord symbols demonstrates a greater level of disparity from the hand annotated key sequences than the triad based models. A weakness in *BSch7* is that it is weighted towards major key outputs due to the dissimilarity between the number of chord symbols represented per major or minor mode, (21 chord symbols for major, and 30 chord symbols for minor, see section 2.2.1). The larger number of chords representing the minor key compared to the chord quantity defined for the major mode reduces the proportional value of individual minor key chord symbols when

Table 4.10: Key detection results for observation matrix BSch7 in conjunction with all four A matrices: error average (Err), distance value for key differences average (Dist), percentage of modulation timing match (Conc), number of modulations as a percentage of hand annotated number of modulations. Ground truth MIDI and transcribed file sets.

A Matrix ↓	*Err*	*Dist*	*Conc*	*Mods*
MIDI				
ANeutral	0.34	0.47	18.93	70.13
AKrum	0.35	0.47	23.24	110.36
ASchEq	0.34	0.47	27.12	113.27
ASchNonLin	0.36	0.47	21.44	109.61
Transcribed				
ANeutral	0.36	0.57	21.65	153.22
AKrum	0.35	0.50	34.66	205.22
ASchEq	0.36	0.51	37.02	238.45
ASchNonLin	0.37	0.49	29.14	217.29

the data is normalised to sum to 1. (I.e. during the process of normalisation the chord values are divided by the sum of chord values to ensure that the rows sum to 1; the greater the total sum of the row, the smaller are the individual chord values following normalisation.) This pushes the outputs of the models over to the major key. For example, chorale BWV 85.6 oscillates between E♭ major and C minor, being in the minor key for a little less than 50% of the time. The four sevenths models detect the minor key for an average of 13% for the MIDI data and 25% for the transcribed audio; the triadic models outperform the sevenths models in the detection of C minor, achieving an average minor key presence of 37% for MIDI data and 35% for the transcribed audio across all triadic models, and therefore being much closer to the actual minor key presence in this chorale.

The three minor key chorales in the set, BWV 85.6, 2.6, and 40.6 have the greatest number of errors for both sets of data, transcribed audio and MIDI. These three chorales were also the least facile to hand annotate due to the fluctuating and inconclusive nature of the harmony. Chorale 40.6, 'Schwing dich auf zu deinem Gott', which is listed as having the most errors for all model versions for both types of data is harmonically

quite equivocal. The opening five chords are ambivalent, and could be interpreted as being either in D minor or F major. Each and every chord is a legitimate member of both keys, and there isn't a decisive indication of either key. It could be argued that the move to an F major chord on the first beat of the second bar, followed by the cadence onto C major, weights the whole phrase towards F major, but it is a first inversion chord, and so still feels a weak inference. D minor is chosen for the hand annotation based on the first chord which is a D minor chord, and because the phrase as a whole has a D minor feel. It is not unacceptable that all of the eight HMM sequences generated by the sevenths HMMs commence with a clear series of F major values. Beyond the first two phrases of chorale 40.6, the four different key transition matrices evidence a variation of output across the file sets, both in terms of key choice and the timing of key change. The chorale moves away from the home key and cadences onto an E major chord in the second half of bar 6 creating an interrupted cadence in A minor. It is anticipated that interrupted cadences could cause incorrect key outputs for the HMM, in fact all of the outputs excepting the Neutral matrix, move correctly to the dominant minor, excepting the Neutral matrix which mistakenly moves to E major. The transcribed data produces the closest outputs, moving to A minor at the beginning of bar 6, whereas the symbolic MIDI data delays the change of key until two beats later, the actual cadence point.

The chorales of less complex harmony, i.e. those which are in a major key and which hardly deviate from this key, result in key sequence outputs which are very similar to each other and to the hand annotated ground truth output for all four models. The clearest example is chorale 1.6, which is quite solidly in F major throughout, resulting in highly consistent outputs across all version of the model. All four models based on the MIDI data recognise the momentary move to the secondary dominant in the final bar; the deficiency of an HMM at phrase boundaries is in evidence here, as none of the output sequences move back to the tonic for the final chord, but remain in C major. The four suspension points in the chorale, all of which contain a G and F a tone apart which cause a 'no chord' value in the ground truth data, are captured with reasonable

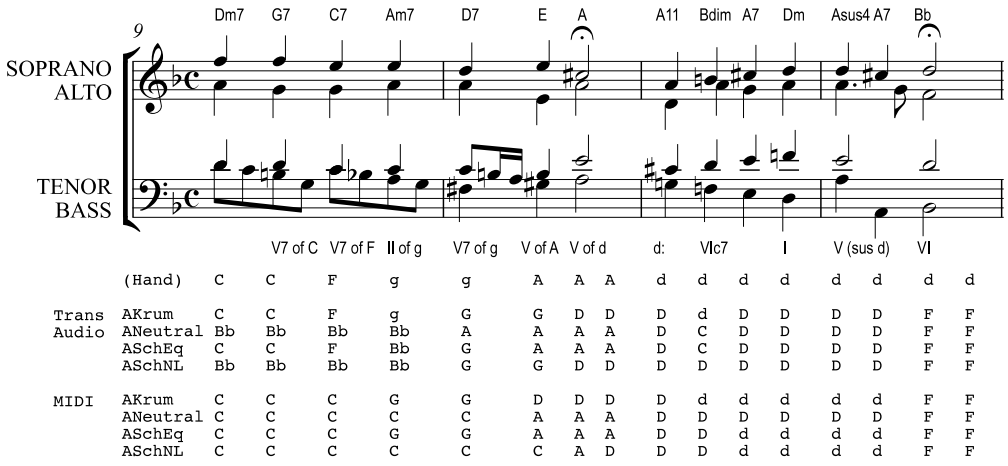


Figure 4.7: Middle bars of BWV 40.6 'Schwing dich auf zu deinem Gott' with HMM key outputs per transition matrix for *BSch7*, hand annotated key and harmony labels using Roman numerals and chord tabs.

accuracy by the transcribed data as C major triads, the two pitches being too close together to extract separately. In contrast, the fragmentation of outputs from the different models reveal areas of complex harmony within individual chorales. Bars 9 to bar 12 of BWV 40.6, (chords 33-48), are particularly inconclusive with regard to key; the harmony is constantly moving and there is sense of flux and ambiguity. At chord 33, shown in Figure 4.7, six of the eight outputs recognise the move to C major for two chords, but from this point the outputs diverge. The transcribed data performs better than the MIDI data, staying in C for the two chords for two of the matrices, then moving on, almost completely correctly for the Krumhansl matrix, to a momentary G minor, before changing key again. The divergence of the outputs speaks of the harmonic intricacy.

The outputs for all file sets for all matrix combinations were ordered per file error rate and distance value, resulting in a highly consistent ordering of the chorales across all of the models, three of which are shown in Table 4.11. The chorales of less complex harmony, i.e. those which are in a major key and which hardly deviate from this key, appear at or near the top of the list, with BWV 1.6 (in the key of F major throughout), disclosing the least errors for almost every model. The three minor key chorales in the

Table 4.11: Chorales ordered by error rate using transcribed audio and *Sch7* models.

	ASchbEq / BSch7			**AKrum / BSch7**			**ANeutral / BSch7**		
	BWV	*Err*	*Dist*	**BWV**	*Err*	*Dist*	**BWV**	*Err*	*Dist*
1	**1.6**	0.18	0.20	**1.6**	0.09	0.09	**1.6**	0.11	0.11
2	**414**	0.20	0.25	**414**	0.20	0.28	**414**	0.23	0.32
3	**253**	0.23	0.70	**140.7**	0.21	0.23	**359**	0.27	0.50
4	**436**	0.25	0.27	**253**	0.23	0.70	**360**	0.28	0.38
5	**140.7**	0.27	0.29	**360**	0.23	0.30	**140.7**	0.29	0.31
6	**360**	0.33	0.44	**436**	0.27	0.30	**436**	0.33	0.36
7	**359**	0.34	0.39	**359**	0.36	0.50	**253**	0.38	0.78
8	**57.8**	0.35	0.46	**57.8**	0.39	0.50	**271**	0.41	0.80
9	**271**	0.42	0.88	**271**	0.42	0.77	**57.8**	0.44	0.87
10	**85.6**	0.45	0.46	**85.6**	0.45	0.48	**2.6**	0.45	0.60
11	**2.6**	0.55	0.67	**2.6**	0.60	0.67	**85.6**	0.48	0.66
12	**40.6**	0.78	1.08	**40.6**	0.80	1.14	**40.6**	0.69	1.16

file set, BWV 85.6, 2.6, and 40.6 are listed at the bottom of the table for all of the models shown.

Analysis of the hand annotated data for the three minor key chorales reveals an average minor key presence of 74%. Comparing the performance of the sevenths models to the triadic models in the detection of the minor key, the four sevenths models average minor key outputs of 24% for the MIDI data and 30% for the transcribed audio. The sixteen triadic models average a minor key output of 52% for MIDI data and 42% for transcribed audio. The triadic models therefore show improved accuracy over the sevenths models, but none of the models achieve the minor key presence of the ground truth for those pieces, suggesting that accurately capturing the minor key is a difficult problem, particularly due to the practice of using a 'tierce de picardié' at cadence points. A possible solution for models containing complex chord types could be to separate out melodic minor and and harmonic minor chord designations, thus defining a much more balanced quantity of chord symbols per key type, but modelling the melodic minor will always generate additional chords due to the changing sixth and seventh degrees in accordance with voice-leading movement. As can be seen in Table 4.12, the chorales with minor key

Table 4.12: Analysis of hand annotated key and chord data to see the relationship between key types and chord distributions.

BWV	No. of Major Keys	No. of Minor Keys	No. of Unique Chord Types
1.6	2	0	13
2.6	1	3	17
40.6	2	2	24
57.8	1	1	11
85.6	1	1	21
140.7	3	1	23
253	3	1	10
271	2	2	22
359	2	1	16
360	2	1	16
414	3	1	11
436	2	1	13

sections feature a greater range of unique chord types than the chorales that are predominantly in major keys. In this corpus, the chord of the diminished seventh appears to be particularly associated with minor keys, an observation which could be modelled in an HMM. (This may not be generally the case for other musical styles.) A possible interpretation of the *BSch7* model is that the results substantiate the notion that triads are more indicative of key than complex chords, excepting the dominant 7th. For this model, the error rates for the transcribed data are very close to the MIDI data achieving a relative best accuracy of 97%.

4.6 Functional Harmony

To produce functional harmony labels for the chorales the automatically labelled chord sequences are combined with the HMM key sequences. The method is to combine the chord symbol and corresponding HMM output key to select the analogous functional chord label for the key from a lookup dictionary. An example of the key:chord label in the dictionary for the pitch class set {5,9,0}, (an F major triad), are shown as follows.

```
Chord:  {5,9,0}
C Major:IV
C♯ Major:III♯3
D Major:III♭R♭5
E♭ Major:II♯3
E Major:IIN6
F Major:I
F♯ Major:VII♯3♯5
G Major:VII♭R
A♭ Major:VI♯3
A Major:VI♭R♭5
B♭ Major:V
C Minor:IV
D Minor:III
E Minor:IIN6
G Minor:VII
A Minor:VI
B♭ Minor:V
```

In the event that a particular pitch class set is not listed as being a chord member of the corresponding key value given by the HMM, a 'nf' (not found) label is returned. An 'nf' can result from either a chord error or a key error. Two of the chorales from the corpus chosen have excerpts in Piston along with his functional harmony labels [Piston, 1983]. These are used for comparison.

A caveat is that any precise and systematic measures of correctness of harmony labelling needs to be considered in light of the anomalous nature of the field. There is rarely a single correct harmony labelling; it is both subjective and sometimes equivocal. The Piston excerpts demonstrate this. Figure 4.8 shows an excerpt from BWV 360, *Werde munter mein Gemute*, along with Piston's functional harmony labels, and the harmony labels generated for both transcribed audio and MIDI data by combining the automatically detected chord labels with the key sequences generated

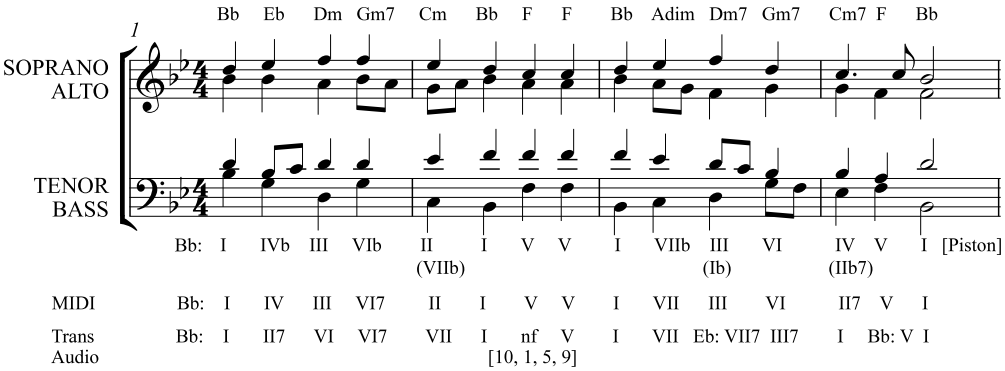


Figure 4.8: Functional harmony labels obtained from MIDI and transcribed audio in conjunction with *BSch7* observation data for BWV 360 'Werde munter mein Gemute' and Piston harmony labels [Piston, 1983], and chord symbols (ours). 'nf' and transcribed chord anomalies result from transcription errors.

by the *BSch7, ASchEq* model. The automatically generated harmony labels do not include inversion labels but these could be obtained with relative ease by using information about the bass note.

As shown in the excerpt Piston offers dual interpretations for three out of the fifteen chords. For example, the chord on the first beat of the final bar is listed by Piston as either a chord IV or a chord II of B♭ major. The chord II is C, E♭, G, and the chord IV is E♭, G, B♭. The segment in question contains E♭, G, B♭ and C of almost equal duration, consequently it could be interpreted as either chord. Although the bass note possibly indicates a leaning towards the chord IV, the chord II with a 7th, considering the strength of emphasis of the C, is an equally valid designation. The chord sequence obtained from processing the MIDI file allocates chord II7 of B♭ at this point, but the transcribed audio version, which modulated two beats previously to the key of E♭ major, and due to a transcription misdetection of the upper C♮, allocates a chord I (E♭, G, B♭). The transcribed audio version also features a pitch transcription error on the third beat of the second bar, producing a chord label that is unrelated to the key of B♭ major, and therefore resulting in an 'nf' label for this chord. For both ground truth and transcribed data, our harmonic labelling results match Piston's harmony labels closely, with the MIDI data matching 13 out of the 15 labels and generating 2 differences of relative insignificance. The

transcribed audio varies more due to transcription errors and modulation in the key sequences, but still produces 6 exact label matches.

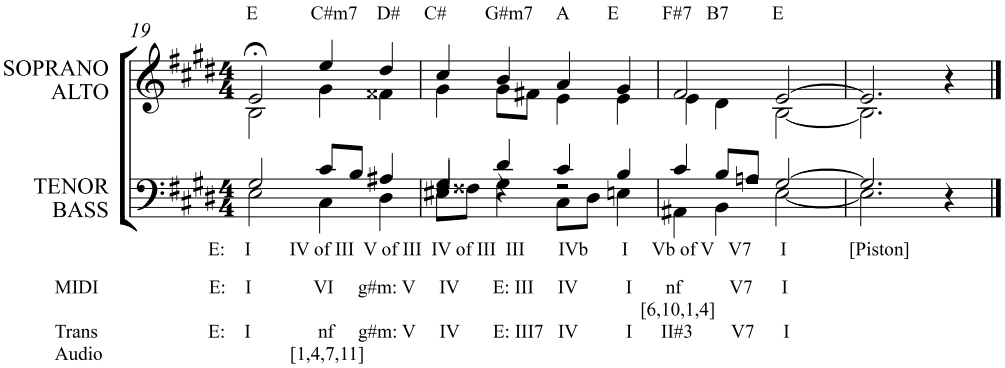


Figure 4.9: Closing bars of BWV 436 'Wie schön leuchtet der Morgenstern' with functional harmony labels derived from *BSch7* observation data in conjunction with *ASchEq* with analysis from [Piston, 1983]

The second excerpt from the data set analysed by Piston is the chorale *Wie schön leuchtet der Morgenstern* shown in Figure 4.9. Piston has assigned labels in the style of secondary dominants. The excerpt shows the closing bars of the chorale, and commences with the terminating chord of the previous phrase, which cadenced in E major.

The previously discussed area of uncertainty when measuring harmony, that of knowing precisely where to mark a change of key, becomes evident with this excerpt. The chord commencing the final phrase is a C♯ minor chord. The chord could equally be a chord VI in E major, or as it is analysed by Piston, a chord IV of G♯ minor. It has been chosen as a pivot chord, a chord of dual function, to smooth the transition from the previous E major phrase into the current G♯ minor phrase. It is not clear that this is where the harmony is taking us, until the following chord, a chord V of G♯ minor. (It is designated V of III by Piston, III being G♯ minor). A simple count based statistic marks this as an error because the model does not change key with the C♯ minor chord, but in the beat following, on the chord V. Piston identifies the first three chords of the phrase as being in G♯ minor, whereas our models identify two. The changeover back to E is a precise match.

The harmony results from the MIDI file demonstrate the mismatch

issue with the functional harmony method as it stands. The chord designation is correctly aligned to the chord [6,10,1,4], the V7 of B major (i.e. V of V - demonstrating the secondary dominant principle being discussed by Piston), but the key output of the HMM at this point is still E major, which is also correct overall. When the key is combined with the chord label, this particular chord is not listed as being a member of the key of E major, and so an 'nf' label is returned. The transcribed audio produces a chord label error for the third beat of bar 19, [1,4,7,11] by labelling all notes within the beat as a single chord. This is not classified as a member of E major, therefore producing an 'nf' at this point. The problematic chord for the MIDI data is circumvented by the transcription process; the closeness in pitch of the F♯ and the E result in a single pitch being chosen here. The chord designation of a chromatic chord II [F♯, A♯, C♯], is not entirely accurate, nonetheless it is a close match.

4.7 Discussion and Conclusions

This chapter has presented an approach to key detection and key modulation using automatic chord classification of transcribed audio and symbolic MIDI data. A set of HMMs were explored using observation and transition probabilities derived from perceptual data and values calculated to represent formal music theory respectively. Although the transcription error rate is quite high, key error rates for the audio recordings are only slightly higher than the key error rates for the ground-truth MIDI. Also, the key error rates are slightly higher for transcribed data using the triadic models, but the complex chord HMM exhibits alignment of results between transcribed audio and MIDI data. This could be interpreted as suggesting that the quality of the transcribed chorales is of sufficiently high quality for the task, however, given that the overall accuracy of the model is relatively low, a greater disparity of results between the data types could emerge should the model accuracy be improved. The question remains open for future experimentation. The music theory models are shown to outperform the perceptual data, with much of the variation

between the models evincing the subtle and often ambiguous nature of musical harmony. Alignment of key boundaries is low overall with the HMM due to the absence of phrase information. The music theory observation matrix *BSchCh* shows the best result for key change concurrence and the music theory matrices demonstrate significant improvement over the perceptual data matrices in this respect. Results are considered promising for the use of automatic transcription research in computational musicology. By combing key outputs with chord sequences, functional harmony labels were obtained for the chorales, opening up opportunities to automatically access information about underlying formal harmonic structures. The methods are promising for the modelling musical style based on higher level abstractions founded in core harmonic theory, for example, measures of modulatory frequency, style of modulation, modulatory sequences and complexity, chord distributions, chord progressions and measures of relative key distance.

In chapter 6 we improve the automatic chord recognition method to be able to classify complex chords and tone groups containing non-chord tones by identifying structural tones. Prior knowledge of key and harmony could also be used to improve the output of a transcription process; for example, initially transcribing the data, obtaining harmony information, and subsequently re-transcribing the data utilising this knowledge. For music research the combination of transcription and musicological models could facilitate the analysis of large corpuses of audio data with the potential for some exciting discoveries about music.

Chapter 5

Creating Ground Truth and MIDI Datasets: The First Twenty-Four Preludes of J. S. Bach's Well Tempered Clavier

> for the profit and use of musical youth desirous of learning, and especially for the pastime of those already skilled in this study [1]

To be able to measure the effectiveness of the automatic chord recognition method presented in chapter 6, a hand-annotated ground truth dataset is required, against which the computational output can be compared. One of the contributions made by the author of this doctorate has therefore been the creation of a ground truth hand-annotated harmony dataset of a test corpus of elaborated keyboard music. The test corpus chosen is the first twenty-four preludes of J. S. Bach's Well Tempered Clavier, Book One. Creating a reference dataset is a laborious task, requiring careful consideration of multiple, often conflicting factors. This chapter relates background information about the test corpus and outlines some of the reasons why the corpus is of such crucial historical and musical significance. The principles guiding the hand annotation process are explained, and the complexities of the harmonic interpretation of ornamental music are discussed. The penultimate section of the chapter

[1] Original title page inscription.

describes the annotation syntax used and the additions that were made to the syntax in order to accommodate the requirements of western harmonic analysis. Preliminary statistical distributions of the hand transcribed data are given at the end of the chapter.

5.1 Historical Context

As attested by the inscription on the original title page, Bach's renowned musical work, the 48 Preludes and Fugues of the Well Tempered Clavier, are acknowledged by scholars to be intentionally pedagogic [Ledbetter, 2002, Kirkpatrick, 1984, Tomita, 2007a]. Historical sources evidence that Bach's approach to imparting his musical prowess was to teach by example; Ernst Ludwig Gerber reports that Bach's lessons to his father involved Bach playing the Well Tempered Clavier several times to him (see Chapter 1 in [Kirkpatrick, 1984]). The collection in Book I dates from around 1720 and was revised a further three times, the latest revision being dated approximately 1736, thus overlapping the creation of Book II chronologically. Neither Book I nor Book II was printed during Bach's lifetime, (the first printed edition appeared in 1799, published by Kollman, London), and due to the differences between revisions and the practice of hand copying, some of which was performed by Bach's students, and much by his wife, Anna Magdalena Bach [Ledbetter, 2002, Tomita, 2007b], different versions of the works exist. Despite a waning of popularity of Bach's music in the one hundred years following Bach's death, a great many more printed editions of the Well Tempered Clavier were published [Palisca, 1981]. The edition on which this research is based is the Associated Board edition of the Well Tempered Clavier, edited by Donald Frances Tovey [Tovey and Samuel, 1924].

5.2 Tuning and Key Integrity

Book I of J. S. Bach's Well Tempered Clavier constitutes the very first complete collection of composed works to use every key as a tonic. The keys progress up the scale chromatically from C, (a departure from the

more accepted ordering around the circle of fifths), alternating between major and minor, and concluding with Prelude 24 in B Minor.

The innovation of the collection is that it was written at a time when modal composition was still the norm, and when instrumental tuning was in a state of experimentation. The modes are rooted in medieval music, and are linked to unequal tuning systems which were believed to give each mode an individual character of its own [Ledbetter, 2002]. A problem is that the number of keys available in unequal tuning systems are restricted to ones with simple key signatures (one to two accidentals), albeit with a greater quantity of perfect intervals [Montagu, 2012]. Three main types of tuning were in use at the time; ¼-comma meantime, Werckmeister III, and an approximation of equal temperament. It is not known what type of tuning Bach used, but the nature of the collection has led scholars to surmise that the tuning system he adopted must have been a version of equal temperament [Ledbetter, 2002]. An important issue of the day was whether equal temperament, a tuning system in which the only acoustically perfect interval is the octave, (the remaining intervals are divided into equal semitones), was an improvement over the traditional modal system. If equal temperament meant simply the possibility of two keys, one major and one minor, that could be transposed around the keyboard, but without each individual key having a distinct and authentic character, then for many of Bach's contemporaries, equal temperament represented an impoverishment of tonal language compared to the expressive possibilities and tonal range of the modes (please refer to the Mattheson and Buttstet debate [Schulenberg, 2006]). In contrast to this viewpoint, the discrepancies between pure overtone intervals and the intervals of tuned instruments, and the variety of responses to this [Lindley, 2009], supports the possibility expounded by some Bach analysts (e.g. [Riemann, 1890]), that each key in equal temperament may still have a unique expressive character. Bach's treatment of particular keys, in terms of expressive style and also types of modulation and key relationships, is thought to be highly influential on the works of later composers, creating precise associations manifested in their compositions (see e.g. [Young, 1991]). The topic continues to be debated, in particular with reference to the implications

of original sources containing versions of pieces in different keys [Tomita, 1990].

5.3 Bach Harmony and Chords

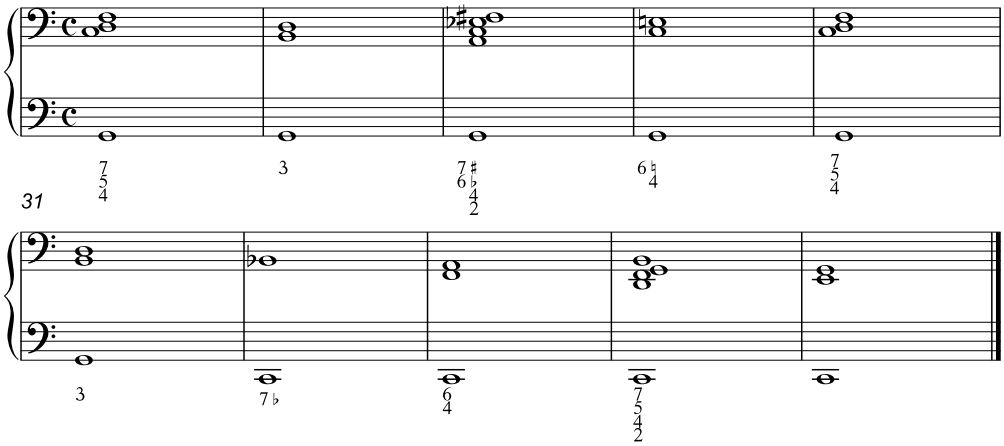


Figure 5.1: Ledbetter's bass figures for the concluding bars of Prelude 1 in C Major, BWV 846, and their implied tones. ([Ledbetter, 2002])

Although it is anachronistic to discuss the music of Bach in terms of Rameau-based theories of harmony ([Rameau, 1971 (Republished]), Baroque music scholars appear to be universally agreed on the idea that 'Bach's melodies are apt to combine in counterpoint so as to form masses of harmony' [Tovey and Samuel, 1924]. Nonetheless there are differences of perspective about whether harmony is the result of melodic movement or whether melodic movement arises from an underlying chord structure. Kirkpatrick states that Bach's harmony is firmly founded in 'the language of thoroughbass' [Kirkpatrick, 1984], page 90. (Thoroughbass and figured bass can be thought of as being synonymous - see chapter 2.) The implication is that Bach's harmony should be thought of in terms of vertical interval aggregates from the bass. Bach would almost certainly have thought of some of these intervals (e.g. unisons, 3rds, 5ths, and 6ths) as being consonant, whereas others (e.g. 2nds and 7ths) would be considered dissonant. Due to the chronological inconsistency between theory

and practice, some music scholars adhere strictly to using figured bass to describe Baroque harmony, listing, for example, a 5-3 or 6-3 sonority to describe a vertical arrangement of tones in accordance with their intervallic distance from the bass note. For example see Figure 5.1, which shows Ledbetter's bass figures for the last ten bars of the C Major Prelude. The stave above shows the tones signified by the figures (my addition), and thus the resulting chords. Even Ledbetter however, opts to use Roman numerals as a convenient means by which to describe harmonic sequences evidenced in the preludes (see *Section 3*, *Preludes*, *Book I*, *The Invention Principle* in [Ledbetter, 2002]).

Vertical combinations of tones could otherwise be termed, for example, a triad or (using Rameau's inversion theory), a first inversion of a triad. In later nineteenth and twentieth century theoretic treatises vertical combinations came to be signified by the Roman numeral notation of functional harmony, a theory first expounded by Hugo Riemann in Vereinfachte Harmonielehre (1893), in which each diatonic chord is deemed to have either a tonic, dominant, or subdominant *function* within the key.

If we consider chords to be an 'aggregation of tones' [Hindemith, 1942], there are as many different types of chords as there are combinations of tones, but some are more recognised in music theory, and used in practice, than others. Baroque harmonic theory ends chord recognition at the seventh, chiefly due to writings which relate chord tones to the harmonic series (e.g. [Rameau, 1971 (Republished]). It is a question contended by Schönberg in [Schönberg, 1922], who cites the tendency of theorists to consign vertical tones beyond the 7th to 'accidental harmony'. What precisely is meant by this maxim? That the harmony has no harmonic significance and is merely an accident of voice-leading? That the composer did not intend the dissonant effect of melodic writing if the dissonance involved is a 9th, 11th or 13th? Despite the theoretic argument surrounding Baroque chord recognition and labelling, sonorities emphasising the vertical intervals of the 9th, 11th, and 13th above the degree of the scale are present in Baroque repertoire. Different scholars adopt different approaches to designating these occurrences, generally in line with their own particular thinking on the topic. For example consider Bach's Partita No. 5, shown

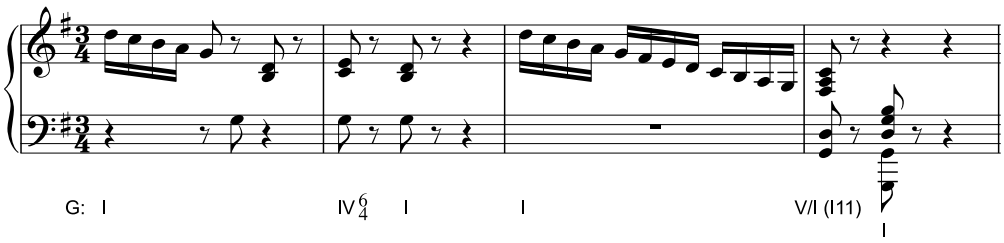


Figure 5.2: Example of Piston's labelling of a tonic eleventh in the *Preambulum* of Bach's Partita No. 5. ([Piston, 1983])

in Figure 5.2, which shows Walter Piston's use of Roman numeral notation along with figures to label the dissonance of a tonic eleventh chord (bar 4) [Piston, 1983]. It seems impossible that the eleventh sonority, occurring as it does on the first beat of the bar in the form of a five note homophonic chord, was either of harmonic insignificance, or unintended by the composer.

For the purposes of musical style modelling, it is mooted that all of the subtle nuances of vertical sonorities present in the corpus need to be captured in order to provide insight into the character of the composer's harmonic texture, particularly those situations presenting unusual note combinations, outside of the common chords. Consequently, despite the contention between rival harmonic theories in relation to Baroque harmony, sonorities featuring extended dissonance in this corpus are labelled as extended chords.

The principles used to create the hand-annotated ground truth chord data from the corpus are reviewed in more depth in the following sections.

5.4 The Preludes

For the purposes of providing a core dataset for the study of voicing, chords and harmony, the preludes alone have been selected. Although preludes and fugues exploit both linear (melodic) and chordal techniques, the preludes can be interpreted as being based more on elaborated harmonic structures whereas the fugues exploit linear or melodic processes.

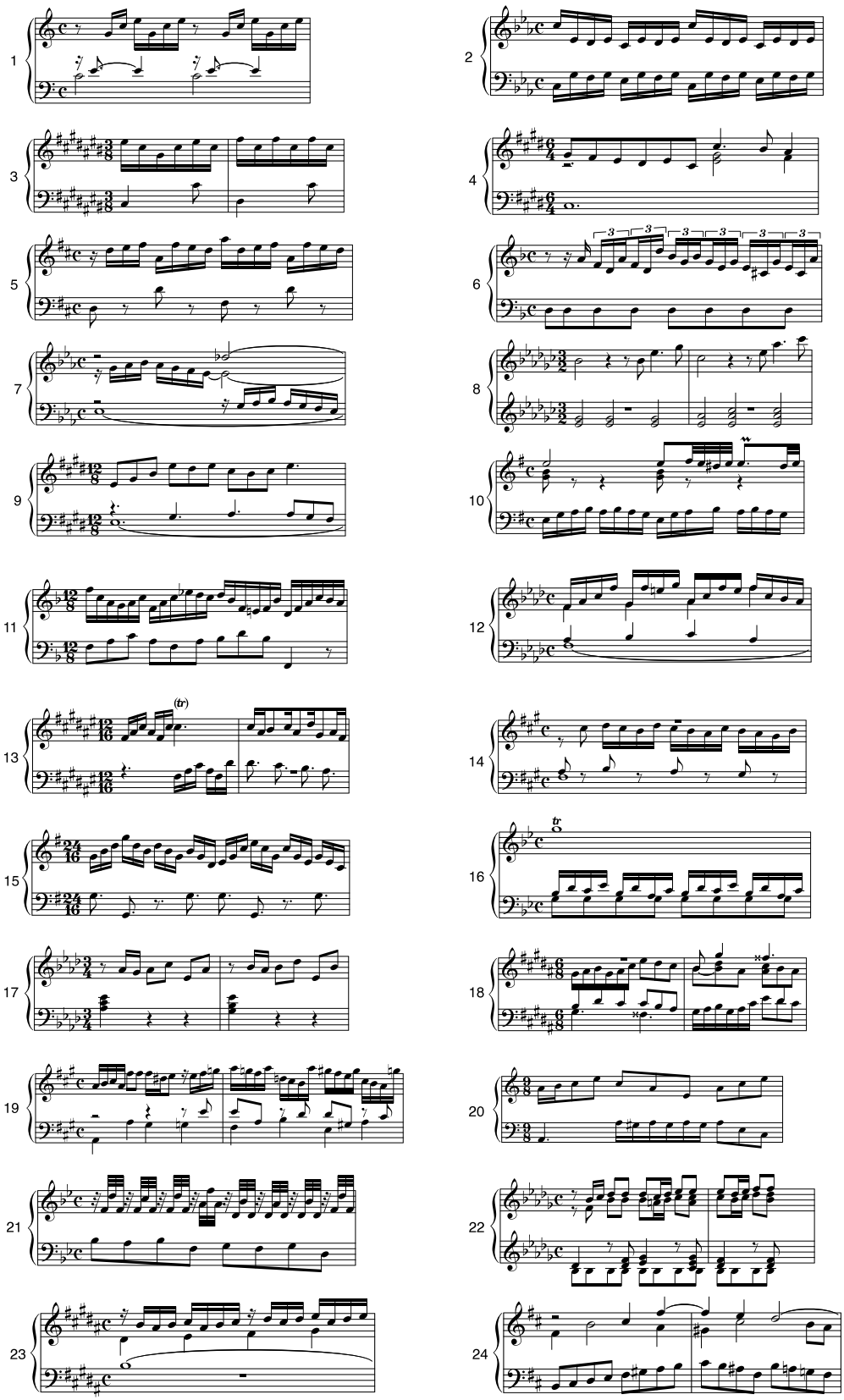


Figure 5.3: Preludes 1 - 24, opening bars.

There is evidence to suggest that the preludes and fugues were originally two separate collections, being only conjoined later on, and therefore the extent to the which each prelude and fugue is a pair is subject to debate [Ledbetter, 2002].

The preludes, the opening bars of which are listed in Figure 5.3, exhibit considerable variety of compositional technique and can be as much a rich resource for MIR research as they have long been a source of study for musicians. The term prelude refers to a piece of music designed to be played as an introduction (for example, preceding a fugue), although this association no longer holds for the piano pieces of later composers such as Debussy. The character of the prelude has also evolved over the ages, referring in its earliest inception (15th and 16th Centuries) to keyboard works particularly noted for their freedom of technique from strict contrapuntal methods due to their basis in improvisation [Ledbetter and Ferguson, 2012]. Preludes, from the earliest organ improvisations up to the modern day, are particularly stylised by idiomatic keyboard writing and freedom of form. Several forms are mentioned in association with the first twenty-four preludes: the toccata, a keyboard composition in free idiomatic style employing full chords and running passages which may contain sections of imitation; three voiced trio sonatas, concerto style, meaning works of three sections; and inventions, a term used by Niedt to refer to works elaborating harmonic structures.

The preludes exemplify the usage of almost every symmetric metre, as can be seen from the list of time signatures in Table 5.1. The concepts of metre and beat, and the distinction between compound, simple, triple and duple time signatures are explained in section 2.4.

5.5 The Annotations

> Bach's melodies are apt to combine in counterpoint so as to form masses of harmony [Tovey and Samuel, 1924]

Despite the wealth of publications about Bach's Well Tempered Clavier, no complete sets of harmonisations exist. The most extensive harmonic

Table 5.1: Summary of Main Characteristics of J. S. Bach's Well Tempered Clavier, Book One, Preludes 1 - 24.

Prelude	BWV	Key	TimeSig	Sig Type	Beat Value	Average Harmonic Rhythm	Characteristics
1	846	C Maj	$\frac{4}{4}$	Simple Quadruple	♩	𝅝	Alla breve, arpeggiated.
2	847	C Min	$\frac{4}{4}$	Simple Quadruple	♩	𝅝, ♩ in Presto	Technical rapid semiquaver movement.
3	848	C♯ Maj	$\frac{3}{8}$	Simple Triple	♪	♩.	Light, ornamental and melodic.
4	849	C♯ Min	$\frac{6}{4}$	Compound Duple	𝅗𝅥.	𝅗𝅥	Rich and weighty minor key, chordal.
5	850	D Maj	$\frac{4}{4}$	Simple Quadruple	♩	♩	Compound melody implying a four voice texture.
6	851	D Min	$\frac{4}{4}$	Simple Quadruple	♩	♩	Triplet semiquaver activity, chromatic progressions.
7	852	E♭ Maj	$\frac{4}{4}$	Simple Quadruple	♩	♩	Sectional, pedal notes, four part double. fugue from bar 25.
8	853	E♭ Min	$\frac{3}{2}$	Simple Triple	𝅗𝅥	𝅝.	Declamatory chordal texture with singing melody.
9	854	E Maj	$\frac{12}{8}$	Compound Quadruple	♩.	♩. , varying to 𝅗𝅥.	Light arpeggiated triplets.
10	855	E Min	$\frac{4}{4}$	Simple Quadruple	♩	𝅗𝅥	Highly decorated.
11	856	F Maj	$\frac{12}{8}$	Compound Quadruple	♩.	♩.	Ornamental arpeggiation of triplets, frequent lengthy trills.
12	857	F Min	$\frac{4}{4}$	Simple Quadruple	♩	♩	Four part compound melody, some use of pedal tones.
13	858	F♯ Maj	$\frac{12}{16}$	Compound Quadruple	𝅘𝅥𝅯.	Varying	Original time signature was 12/8 features syncopated. demisemiquavers
14	859	F♯ Min	$\frac{4}{4}$	Simple Quadruple	♩	♩	Technical prelude, Bach notably used the thumb on raised notes.
15	860	G Maj	$\frac{24}{16}$	Compound Octuple	𝅘𝅥𝅯.	Varying	Arpeggiated semiquaver triplet movement.
16	861	G Min	$\frac{4}{4}$	Simple Quadruple	♩	Interpretive	Highly ornamental, trills, demisemiquaver decorative style.
17	862	A♭ Maj	$\frac{3}{4}$	Simple Triple	♩	Bar	Varied texture of block chord and rapid figurations.
18	863	G♯ Min	$\frac{6}{8}$	Compound Duple	♩.	Varying from ♪ to ♩.	Example of upper voice pedal at end.
19	864	A Maj	$\frac{4}{4}$	Simple Quadruple	♩	♩	Example of triple counterpoint.
20	865	A Min	$\frac{9}{8}$	Compound Triple	♩.	Varying from bar to ♪.	Light decorative three part texture.
21	866	B♭ Maj	$\frac{4}{4}$	Simple Quadruple	♩	♪	Dense chords and brilliant runs of demisemiquavers.
22	867	B♭ Min	$\frac{4}{4}$	Simple Quadruple	♩	Varying	Very dense chords in places, pedal effect of repeating quavers.
23	868	B Maj	$\frac{4}{4}$	Simple Quadruple	♩	♩	Mixed rhythmic figures in a three part texture.
24	869	B Min	$\frac{4}{4}$	Simple Quadruple	♩	♩	Vocal four part texture.

analyses of the corpus are those published by Riemann in 1890 [Riemann, 1890]. Riemann's idiosyncratic annotation syntax takes a little deciphering (see Figure 5.4), and his analyses are predominantly restricted to short excerpts. They are also unsuitable for systematic processing without a further stage of translation and syntactic adaptation to produce a machine readable format. Despite these limitations Riemann's analyses provide a valuable and authoritative source, and Riemann's judgements are deferred to as the definitive label in the annotations for the sections of the corpus that they are available for. This section of the thesis describes the annotation process, and the challenges and guiding principles adopted.

Annotating the underlying chords of elaborated keyboard works such as the preludes requires the concurrent processing of note, chord, and contextual information. A label is chosen based on the aforementioned Riemann source, and then in accordance with the musical expertise of the annotator about the relative structural significance of the notes and their potential as chord and/or key members. The process involves deciding whether any of the notes have a solely ornamental or melodic role and may therefore be omitted with respect to chord definition. At the same time, the group of notes must be considered with reference to immediate, local, and global context, to arrive at credible and justifiable chord designations. Other texts quoted in this section of the thesis are referred to for guidance about both key and chord, these include [Ledbetter, 2002, Kirkpatrick, 1984, Tomita, 2007a, Schenker and Salzter, 1969, Piston, 1983].

In a contrapuntal texture constituent tones may be missing from any of the chord types, for example a triad may be implied by the presence of a single unison. In the case of complex dissonances, several elements may be absent. The process of chord selection is interpretive and subjective, and on the part of the annotator, higher level information is taken into account. This includes knowledge of key, preceding and succeeding note groups, melodic, registral and thematic arrangements, the formal position of the group of notes within the prelude, and the metrical and durational emphasis of notes. There are many cases where the underlying chord is transparent, about which all musicians will agree, for example much of

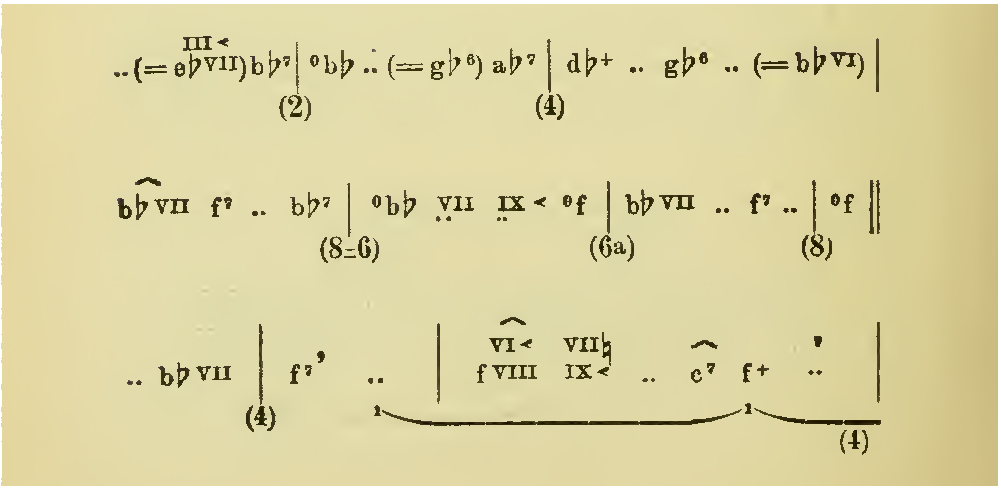


Figure 5.4: A fragment of Riemann's harmonic analysis of the Prelude in B♭ Minor [Riemann, 1890], one of the more lengthy analyses in the publication.

the arpeggiation in Prelude No. 1 in C Major, BWV 846. Even here however, there are differences in the translations made by different musical authorities. For example, bar 23 arpeggiates upwards from A♭ in the bass through F, to the semiquaver figuration of B, C, D (Figure 5.5). Schenker [Schenker and Salzter, 1969] chooses the F Minor triad with added 6th for this bar. Riemann marks this bar as a dominant minor 9th minus the root (G, B, D, F, A♭ where G is the missing root) [Riemann, 1890]. Ledbetter gives two figured bass versions of the prelude; in the earlier version the Ab is absent (i.e. there is a different chord in the earlier version), in the final version Ledbetter's figured bass lists an added fourth (D) and third (C) [Ledbetter, 2002]. The example highlights the dichotomy between the computational requirement for rigorous, accurate and consistent ground truth data, and the inherent elusiveness of tonal harmony. Whether intentional on the part of the composer or not, a central element of musical expression is its exploitation of blurred boundaries and unclear classifications. With all ground truth data about harmony, it is important to acknowledge that there is no one single correct answer, and that musicians and musicologists will vary in their opinion about what is the correct chord or key.

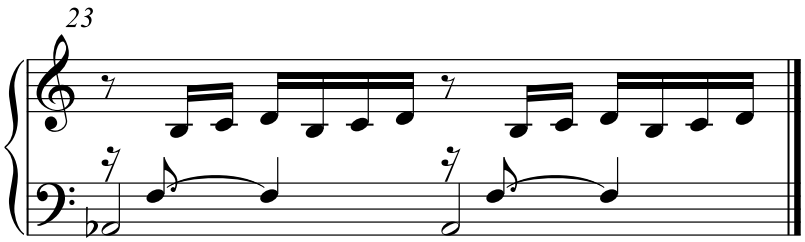


Figure 5.5: Prelude 1 in C Major, BWV 846, Bar 23.

5.5.1 Harmonic Rhythm

To provide a data set of consistent granularity that can be used for audio chord recognition as well as computational musicology, and similarly to annotations provided for popular music [Mauch et al., 2009], a beat synchronous approach was adopted. There is perhaps a sense of incongruity in creating chord labels at consistent and regular time intervals, given that the rate of change of harmony in a piece of music, known as the harmonic rhythm, varies. Harmonic rhythm is independent of melodic rhythm, and generally does not adhere to uniform time intervals. If one views excerpts of harmonic analysis in [Piston, 1983]; it will be immediately apparent that chord labelling alters from bar to bar in terms of duration, frequency, and position.

It may seem that this is not a particular problem. If an entire bar is labelled as a V7 chord (Riemann's labels vary from a single label for a single bar, therefore reflecting the extended chord indicated by the content of the whole bar, or may give a unique label per beat) one might think that the solution is to label each individual beat in the bar with a V7. In fact, the level of focus impacts chord labelling more literally than this. A full or half bar inference of a V7 chord does not mean that smaller beat level groups also imply a V7 chord. A typical example is that of a V7 chord expressed over 2 beats; each beat may contain the notes of a subset of the larger focus chord, for example, a dominant triad followed by the diminished triad on VII. A contrasting situation is when harmonic change occurs at a faster rate than the level of annotation, for example, when two chords are stated successively within a single beat. It is not practical

or realistic to endeavour to represent every harmonic nuance of melodic movement; with all approaches to harmonic analysis, and especially in the case of decorative keyboard music, demarcating precise boundaries is difficult and subjective. All annotators, whether following harmonic rhythm or not, aim to portray the most plausible and suitable set of chord labels for the purpose for which they are intended. The beat synchronous approach therefore results in a singular set of chord labels that may diverge from a harmonic rhythm approach, but harmonic rhythm is itself to some degree open to interpretation, and in general, a change of harmony, whilst not occurring uniformly, mostly occurs in time with a beat.

As mentioned previously, the durational value of a beat is affected by time signature; generally, compound time signatures indicate compound beat values. There is an exception to this in the set. Prelude 3 in C♯ Major, BWV 848, (Figure 5.6), which has a simple triple time signature of $\frac{3}{8}$, has such a sparse texture that this prelude is annotated at the compound beat level.

5.5.2 Harmonic Dualism

The preludes present many situations similar to the aforementioned Bar 23 of BWV 846, in which the tones present in a group are open to multiple possible interpretations. Bach's music regularly signifies several possible chords, and sometimes keys, at once. Consider for example, the excerpt in Figure 5.6, which shows bars 70 to 73 of Prelude 3 in C♯ Major. There is an ascending melodic sequence in the upper voice from E♯, through F double sharp, to G♯ in bar 73. Bar 70 outlines the tonic chord. But how to interpret bars 71 and 72? The left hand is enunciating C♯ Minor, but the enharmonic equivalent of F double sharp is G♮, and the combination of C♯, E♮, G♮ and A♯ (enharmonic B♭), form a diminished seventh chord. The F double sharp, high up in the register and on the first beat of the bar, occupies a level of salience that could overcome the bass G♯.

The excerpt is just one example of many in the corpus which raises questions about the perception of harmony and the link between perception and harmonic interpretation. It could be argued that the enharmonic

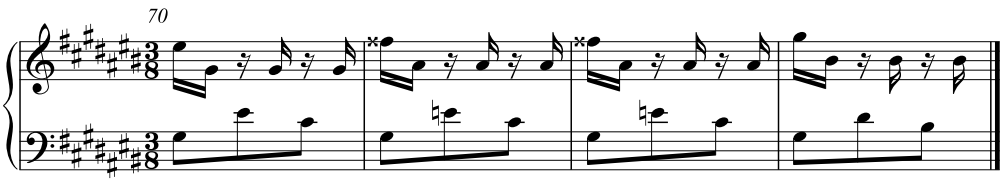


Figure 5.6: Prelude 3 in C♯ Major, BWV 848, Bars 70-73.

notation suggests that the diminished 7th grouping is incidental, but the sound of a diminished 7th chord is so distinctive it is unlikely that the composer was unaware of the effect. What would a listener perceive here? Would they hear two chords played simultaneously - the left hand C♯ Minor pattern and the diminished seventh of the combination? Alternatively, if a single sonority is discerned, would the tonic minor triad be perceptually prior to the dissonance of the diminished 7th? If the diminished 7th is recognised as the overriding harmony, how would a listener account for the G♯; as a pedal tone, or as inessential? It would be interesting, but is beyond the scope of this thesis, to discover how the perceptions of listeners either validate or contradict the annotations.

5.5.3 Dissonance

In keeping with compositional style of the time, Bach's presentation of dissonance is primarily in the form of suspensions (prepared dissonance) and appoggiatura (unprepared dissonance). An example of this can be seen in Figure 5.7, which shows an excerpt from Prelude 7, BWV 852, featuring a series of suspensions. The excerpt opens at bar 31, with a major 9th on Eb {Eb, G, (B♭ - absent), D, F}, where the 7th (D) and 9th (F) are presented in the form of a double suspension, (i.e. held over from previous beats), in the upper voices, which subsequently resolve downwards by step. Bar 32 follows with an established style of elaboration of the period in which all elements of a 9th chord (in this case the 9th on G, {G, Bb, D, F, A♭}), are articulated melodically. The held notes at this point (D and G), are not dissonant, but maintain the overall texture. The 9th is implied across two crotchet beats, (although only actually stated in

the first crotchet beat), and is followed by a minor 7th on C, also for two beats. The two chords highlight one of the difficulties of harmonic rhythm discussed earlier: zooming into the crotchet level, the second beat of of the 9th chord states the chord only up to the 7th, and could also indicate an E♭ major triad, similarly, the second beat of the C minor 7 when viewed in isolation, could also be interpreted as the dominant 7th on F {F, A♮, C, E♭}. The 9th chord on F at the beginning of bar 33 would be marked as a dominant 7th on F by many, and is included to demonstrate the challenge of producing consistent data when interpreting embellished music.

Chords containing an 11th or 13th are often presented during the Baroque period in the form of a pedal note rather than as a clear chord member, which nonetheless creates a dissonant sonority. A technique frequently used by Bach is to superimpose a dominant 7th or 9th chord over the tonic, creating the effect of an 11th or 13th (see Figure 5.8). Dissonances such as these generally resolve correctly by step, e.g. onto tonic harmony. One of the most difficult aspects of annotating any of the dissonant chords, is judging the harmonic importance of the dissonant tones. The ninth of bar 32 in Figure 5.7 is relatively uncontroversial, but there are many cases when the sounding of the upper dissonances are more fleeting, and it is not always clear whether the notes warrant inclusion in the definition of the chord. The principles used to disambiguate this include: affording greater harmonic significance to dominant dissonance; melodic, metrical and registral accentuation of notes; harmonic context; the overall quantity of dissonance, and the level of harmonic ambiguity in the work as a whole. This latter tends to be much greater in the minor key pieces.

5.5.4 Texture

Music written for keyboard instruments which precede the pianoforte require a reasonably intense degree of movement to maintain both musical interest and sound volumes due to the lack of sustaining sound of period instruments. Rapid ornamental movement combined with rhythmic variation often results in partial articulation of harmonies, with too few chord member notes per beat group to conclusively define a specific chord. For

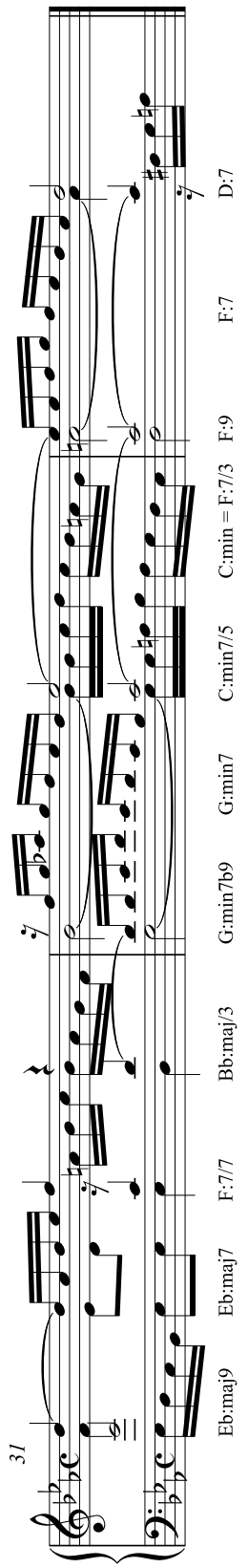


Figure 5.7: Prelude 7 in E♭ Major, BWV 852, Bars 31-33.

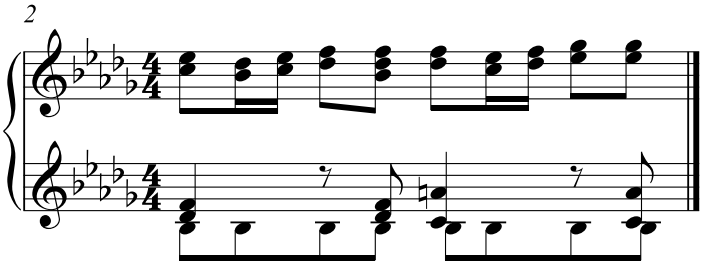


Figure 5.8: 13th effect produced by dominant minor 9th over tonic pedal on the 4th beat of bar 2 of Prelude 22 in B♭ Minor, BWV 867.

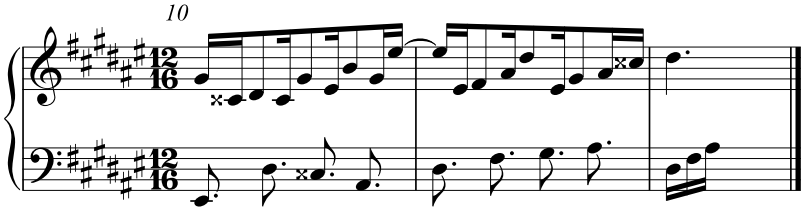


Figure 5.9: Prelude 13 in F♯ Major, BWV 858, Bars 10-12.

example, view Figure 5.9, which shows the syncopated rhythm and sparse, fragmentary texture of Prelude 13 in F♯ Major, BWV 858. The excerpt features a segment of the prelude modulating to D♯ Minor. Bar 10 is problematic to annotate, with the conflicting presence of C double sharp and D♯ in both voices. Despite the presence of D♯, the first two beats imply a C double sharp diminished triad, moving subsequently to the dominant seventh of D♯ Minor {A♯, C double sharp, E♯, G♯}. In the case of such sparse data, knowledge is levied of key, local context, and implied chord progressions.

5.5.5 Chord Inversion

Chord inversion refers to the scale step of the bass note in the vertical arrangement of pitches in a chord (Section 2.2.1). In a moving elaborated texture the chord position is not necessarily held throughout the beat, for example, the bass may travel from the root note to the third. Standard broken chord or returning patterns in the bass which commence on a

specific degree are treated as having the position of the first note of the pattern even if the pattern progresses to a different chord position. When there is much stepwise movement it is not always possible to identify a single chord position, in which case chord position is not annotated and root position is implied.

5.5.6 Pedal Tones

Pedal tones are tones which are maintained across more than one bar [Apel, 1970]. They can appear in any of the voices but are more frequently found in the bass. Pedal tones take the form of sustained or repeated notes, alternating octave patterns or a repeating pitch of shorter durational value at a regular metrical position in the bar for a series of bars. A common use of pedal tones in the preludes is as a tonal anchor during a chord sequence that begins with, departs from, and subsequently rejoins the pedal tone. Pedal tones are demarcated separately in the annotations (see next section).

5.6 Chord Annotations

The annotations make use of Chris Harte's chord syntax [Harte, 2010]. The syntax defines a method of representing chords independent of key context, in which the note members of a chord are unmistakable. The syntax has many advantages; it is clear, unambiguous, and requires little or no musical training to understand. For these reasons it is a popularly utilised syntax which has been widely accepted by the MIR community. Chris Harte's syntax, and his annotations of The Beatles are well known in the MIR community, consequently, it is deemed sufficient to refer the reader to Chris Harte's PhD thesis [Harte, 2010]. (Please also see Table 5.2.)

The manual annotation of keys and chords is a lengthy and time consuming task. Consequently, although the original syntax allows for the explicit labelling of arbitrary sets of chord intervals, and provides some shorthand labels up to and included three types of 9ths, to expedite the

Table 5.2: Chord annotation syntax extending Harte's annotation syntax [Harte, 2010]: showing new chord descriptors, shorthand notation, musical intervals, successive semitone interval content, and note examples. Asterisks denote shorthand labels that were not in Harte's syntax.

Description	**Shorthand**	**Interval List**	**Semitone Intervals**	**Example**	**New**
Major	maj	(1,3,5)	(4,3)	G,B,D	
Minor	min	(1,b3,5)	(3,4)	G,B♭,D	
Diminished	dim	(1,b3,b5)	(3,3)	G,B♭,D♭	
Augmented	aug	(1,3,#5)	(4,4)	G,B,D♯	
Seventh	7	(1,3,5,b7)	(4,3,3)	G,B,D,F	
Major Seventh	maj7	(1,3,5,7)	(4,3,4)	G,B,D,F♯	
Minor Seventh	min7	(1,b3,5,b7)	(3,4,3)	G,B♭,D,F	
Diminished Seventh	dim7	(1,b3,b5,bb7)	(3,3,3)	G,B♭,D♭,F♭	
Half Diminished Seventh	hdim7	(1,b3,b5,b7)	(3,3,4)	G,B♭,D♭,F	
Minor Major Seventh	minmaj7	(1,b3,5,7)	(3,4,4)	G,B♭,D,F♯	
Augmented Seventh	aug7	(1,3,#5,b7)	(4,4,2)	G,B,D♯,F	*
Augmented Major Seventh	augmaj7	(1,3,#5,7)	(4,4,3)	G,B,D♯,F♯	*
Major Sixth	maj6	(1,3,5,6)	(4,3,2)	G,B,D,E	
Minor Sixth	min6	(1,b3,5,6)	(3,4,2)	G,B♭,D,E	
Italian Sixth	It6	(1,3,#6)	(4,6)	G,B,E♯	*
German Sixth	Gr6	(1,3,5,#6)	(4,3,3)	G,B,D,E♯	*
French Sixth	Fr6	(1,3,#4,#6)	(4,2,4)	G,B,C♯,E♯	*
Ninth	9	(1,3,5,b7,9)	(4,3,3,4)	G,B,D,F,A	
Major Ninth	maj9	(1,3,5,7,9)	(4,3,4,3)	G,B,D,F♯,A	
Minor Ninth	min9	(1,b3,5,b7,9)	(3,4,3,4)	G,B♭,D,F,A	
Seventh Minor Ninth	7b9	(1,3,5,b7,b9)	(4,3,3,3)	G,B,D,F,A♭	*
Minor Seventh Minor Ninth	min7b9	(1,b3,5,b7,b9)	(3,4,3,3)	G,B♭,D,F,A♭	*
Eleventh	11	(1,3,5,b7,9,11)	(4,3,3,4,3)	G,B,D,F,A,C	*
Major Eleventh	maj11	(1,3,5,7,9,11)	(4,3,4,3,3)	G,B,D,F♯,A,C	*
Minor Eleventh	min11	(1,b3,5,b7,9,11)	(3,4,3,4,3)	G,B♭,D,F,A,C	*
Halfdim Seventh Eleventh	hdim711	(1,b3,b5,b7,9,11)	(3,3,4,4,3)	G,B♭,D♭,F,A,C	*
Thirteenth	13	(1,3,5,b7,9,11,13)	(4,3,3,4,3,4)	G,B,D,F,A,C,E	*
Major Thirteenth	maj13	(1,3,5,7,9,11,13)	(4,3,4,3,3,4)	G,B,D,F♯,A,C,E	*
Minor Thirteenth	min13	(1,b3,5,b7,9,11,13)	(3,4,3,4,3,4)	G,B♭,D,F,A,C,E	*
Minor Seventh Minor Thirteenth	min7b13	(1,b3,5,b7,9,11,b13)	(3,4,3,4,3,3)	G,B♭,D,F,A,C,E♭	*
Thirteenth Minor Ninth	13b9	(1,3,5,b7,b9,11,13)	(4,3,3,3,4,4)	G,B,D,F,A♭,C,E	*

process of manually annotating the harmony of a more diverse range of musical styles, a new set of shorthand labels are provided to promote the rapid and easy encoding of extended chords. This includes 7th and 9th chords, as well as 11ths, 13ths, and augmented 6ths. Extended chords are a primary distinguishing feature of western musical style period, with post Romantic music, and jazz in particular exploiting the colourful effect of 11th and 13th chords [Piston, 1983, Strunk, 1988], consequently being able to easily and rapidly label such chord instances in the repertoire is extremely important. The shorthand labels shown in Table 5.2 are understood to symbolise a precise pattern of intervals based on a specific scale degree, for example, *C:7b9* refers to the seventh minor ninth chord on C, or *C, E, G, B♭, D♭*. We intentionally adopt jazz style chord notation for the labelling of 'altered' chords in order to produce a chord dictionary that is as generically applicable and as consistently notated as possible whilst continuing to be musically intuitive. The additions listed in Table 5.2 are representative of the dissonances used in classical western harmony, from the Baroque to the Romantic era. Conventionally not all of the tones are sounded together, although they may be touched upon melodically; the 3rd and 5th are typically omitted from 11th chords, and the 5th, 9th and 11th from 13ths. We plan to further expand this standard set of chord definitions in order to create a comprehensive dictionary of chords encompassing broader musical periods and jazz (see future work).

A further addition to the syntax is a method of identifying the presence of a pedal note along with a chord. A pedal note cannot simply be annotated as a bass note below a chord; the pedal note may be completely unrelated to the chord, and may also not be a bass note. In addition, using the bass note part of the syntax to show a pedal note would mean removing the representation of the actual chord bass tone. Pedal tones are therefore indicated by a tilde followed by the pedal tone e.g.

`G:maj/3 ~ C`

The interpretive nature of harmony has been mentioned earlier. The aim whilst annotating the preludes was to select the single most representative chord choice for a beat group, but in some situations more than one possible chord appears to be equally valid. To allow for more accurate

evaluation of automatic algorithms by avoiding arbitrary choices which a system could not be expected to guess correctly, in such cases all possible chord interpretations are annotated. The distinction between chord choices is shown by an equals sign, e.g.

```
B:dim7/3 = D:dim7
```

Note that this notation does not assert the equivalence of the two chord labels, but that they are equally valid descriptors of the harmony (e.g. see Figure 5.7, last beat of bar 32).

5.7 Key Representation

The annotations discussed thus far provide a literal representation of underlying chord sequences in the preludes. There is a limitation imposed by using an annotation syntax tailored to popular music to encode the chord sequences of advanced Baroque keyboard composition. Regardless of notational particulars and the ease or otherwise of adapting these for computational purposes, the critical difference between traditional musicological methods of harmony annotation and the approach used here, is one of scope: literal chord sequences, annotated independent of key context, do not represent the implied harmony.

Harmony is not easy to define succinctly. The term encapsulates many concepts, from chords, consonance and dissonance, and cadences, to key and modulation, to harmonic function, structure and musical form. Harmony refers to the ways in which entities such as scale, chord and key are combined to create musical coherence. It is the skill and variety of these combinations that allow composers to construct musical structures capable of supporting large scale musical works. Dahlhaus [2007] explains that in order to understand harmony, one must be able to relate chords (in this case chords are already understood in terms of their function within a key) in relation to metre, musical phrasing, and form. Dahlhaus gives the example of a cadential chord sequence, I-IV-V-I, (in the key of C Major this would be the chord series C, F, G, C, all major), which cannot be reversed without significantly altering the musical effect, whereas the harmonic outline of I-V-IV-I, (C, G, F, C), is quite usual for an entire

movement. The advantage of the musicological representation is one of abstraction; by symbolising a chord in terms of its function within a key context, or a key in relation to an overall key, the building blocks of harmony, (for example, cadences), can be described. Without this level of symbolic abstraction, the presence of a first inversion D♭ major chord in a C Major sequence, is just that, a D♭ Major first inversion. If one interprets the chord in its key context, the chord is immediately recognisable as a Neapolitan Sixth, a chromatic triad on the flattened supertonic with a unique harmonic function and place in musical history.

Enhancing the richness of annotated information by including key as well as chord, from which functional harmony can be derived, is clearly advantageous for music research. Annotating key information presents its own issues, not least of which is how to select and represent the precise moment of transition from one key to the next, a process known as modulation. A standard method of modulation is to move into the new key via chords common to both keys, resulting in an area of overlapping keys (see [Piston, 1983]). These segments of music contain chords that can be defined as being either in the key being moved from, or the key being moved to. Musicologists use additional information such as phrasing to decide at precisely what moment the old key finishes and the new key begins, although they may also demarcate the overlap. In addition to this, there is the question of how to annotate sections of music which feature rapidly changing key centres, tonal equivocation, or which use more than one key, as in the case of bitonal composition.

In the interests of maximum ease of use, clarity and simplicity, the key annotations of the preludes, (currently only for the first 5 preludes), adopt the same beat synchronous approach and extended syntax to represent key as that used to represent chords. As with the chord annotations, it is permissible to annotate more than one key per beat. The annotations are therefore able to represent areas of key overlap and polytonality e.g.

```
F:maj = C:maj | F:maj = C:maj | C:maj
```

5.8 Validation and Correction of MIDI Score Data and Hand Annotations

5.8.1 Cleaning and Validating the MIDI data set

The author would like to thank Phillip Kirlin of University of Massachusetts Amherst, for providing voiced MIDI recordings of the preludes. These were used as the corresponding digital score data for the research described in the following chapter. The MIDI data was validated against the edition of the score being used for this research ([Tovey and Samuel, 1924]) via a repeated process of loading the files into a music notation software package and manually checking for discrepancies or errors, particularly of timing or pitch. The corrected data was then exported out of the notation package in MIDI format, with the errors corrected. The process is a painstaking and arduous task, sometimes made more so by the typesetting of the software package, and one that had to be repeated many times over. Errors that were missed during the manual checking process were discovered later, via the method detailed in the next subsection, or during the computational processing detailed in the next chapter. Each error in the data was therefore manually corrected in the notation package, and a new version of the MIDI file once again exported. The data cleansing process was a critical part of the methodology of the research described in the next chapter, meaning that the results of the computational processing could be confidently interpreted as a product of the computing algorithm, rather than the result of data error.

5.8.2 Verification of Hand Annotations and Further Checking of the MIDI Data

To verify the quality of the hand-annotated data and to check further for discrepancies between the MIDI data set used for subsequent experiments, and the edition of the score, a computer program compares the pitch class set versions of the hand annotations to the actual pitch classes present in each beat segment. The program produces intersection, difference, and symmetric difference values, between each corresponding hand-annotated

set and the pitch class set of the MIDI data. The difference value is of most interest, as it shows pitch classes present in the hand-annotated set that are not present in the MIDI data, indicating a possible error. The symmetric difference value shows values present in either group but not the other. For example given a hand-annotated set of [7, 10, 1] and a MIDI pitch class set of [10, 4, 7], the difference value is [1], the symmetric difference is [1, 4], and the intersection of the groups is [7, 10].

The values were written to files and were manually checked against the score to discover and correct errors in either dataset, or to verify that the differences are considered to be legitimate; i.e. they are the result of missing chord tones. Taking Prelude 3 as an example, there are many instances of legitimate differences between the hand-annotated pitch class set representation and the actual pitch classes present in any one beat segment. Throughout the prelude there are decorated dyads of a third apart, which are assumed to signify a triad omitting the fifth, and perhaps due to the light texture of the prelude, many of the 7th chords in this prelude also omit the fifth. Missing chord factors, particularly for complex chords, are a significant challenge both to annotation and automatic chord methods. Table 5.3 shows the number of segments with difference values shown as the percentage of the length of sequence per prelude, and the average number of different pitch classes when a different pitch is found.

5.9 The Annotated Dataset

A contribution of this thesis will be to make the dataset available from the Centre for Digital Music data repository for use by the community. In addition to this there will be a software parser which parses the annotations and produces equivalent sequences of pitch class set

5.10 Corpus Distributions

In this section statistics are drawn from the hand-annotated chord data. Tables 5.4 and 5.5 show the distribution of chord types per prelude, expressed as a percentage of the length. Averages are shown for the set of

Table 5.3: Pitch class difference percentage between the hand-annotated data and the MIDI data set, and the average number of different pitch classes, per prelude.

Prelude	**% Segments with Differences**	**Average difference quantity**
Prelude 1	16.43	1.57
Prelude 2	41.45	1.00
Prelude 3	20.19	1.10
Prelude 4	12.82	1.10
Prelude 5	54.29	1.04
Prelude 6	13.46	1.07
Prelude 7	30.71	1.10
Prelude 8	20.00	1.17
Prelude 9	39.58	1.00
Prelude 10	33.54	1.11
Prelude 11	6.94	1.00
Prelude 12	18.18	1.00
Prelude 13	37.50	1.09
Prelude 14	36.46	1.00
Prelude 15	29.61	1.20
Prelude 16	14.47	1.09
Prelude 17	44.70	1.14
Prelude 18	13.79	1.25
Prelude 19	57.29	1.00
Prelude 20	11.90	1.00
Prelude 21	17.50	1.21
Prelude 22	22.92	1.18
Prelude 23	38.16	1.14
Prelude 24	51.06	1.14
Average	28.46	1.11

12 major and 12 minor key pieces, and for the full corpus of twenty-four preludes. It can be seen from the tables that the most common chords are the major and minor triads (interval profile {4, 3} and {3, 4} respectively), and as might be expected, major chords dominate in the major key pieces, and minor chords prevail in the minor. The most common sevenths are the 7th, (often referred to as the dominant 7th), with the interval profile of {4, 3, 3}, followed by the diminished 7th, interval profile of {3, 3, 3}, and the minor 7th, interval profile of {3, 4, 3}. The most common extended chord type is the 7b9, or dominant minor 9th, for both major and minor key pieces. The distribution table, Table 5.5, evidences the selective and limited usage of extended dissonance in the harmonic

language of the corpus. Interestingly, it also appears to show that a singular type of extended dissonance is characteristic of individual preludes, for example the minor 9th in Prelude 4.

Table 5.6 shows the distribution of root pitch scale degrees, disregarding inversion, and expressed relative to the main key to enable a direct comparison of the data. Commencing at pitch class 0 and progressing chromatically up the scale, 0 therefore symbolises the scale degree of the tonic, 1 the semitone above that, 2 the supertonic, 5 the subdominant, and 7 the dominant. In the major key pieces, chords built on the tonic, dominant and supertonic scale degrees are most prominent, whereas in the minor key pieces, chords built on the tonic, dominant, subdominant, and then supertonic, are the most conspicuous.

Table 5.4: Distributions of triads, sixths and sevenths in hand annotated data as a percentage of sequence length per prelude.

Prelude	**maj**	**min**	**dim**	**aug**	**min6**	**maj6**	**7**	**dim7**	**maj7**	**min7**	**halfdim7**	**minmaj7**	**augmaj7**
Major Keys													
1	31.43	8.57	12.14	0.00	2.86	0.00	25.00	2.86	8.57	8.57	0.00	0.00	0.00
3	43.27	21.15	3.85	0.00	0.00	0.00	10.58	10.58	2.88	5.77	0.00	0.00	0.00
5	29.29	18.57	5.71	0.00	0.00	0.00	26.43	11.43	0.71	6.43	0.00	0.00	0.00
7	41.43	20.71	6.79	0.36	0.00	0.36	16.07	3.21	3.57	5.36	0.36	0.00	0.00
9	47.92	23.96	5.21	1.04	0.00	0.00	19.79	2.08	1.04	4.17	0.00	0.00	0.00
11	23.61	13.89	15.28	0.00	0.00	0.00	30.56	1.39	2.78	9.72	0.00	0.00	0.00
13	39.17	30.00	11.67	0.00	0.00	0.00	11.67	4.17	0.00	1.67	0.83	0.00	0.00
15	48.03	22.37	10.53	0.66	0.00	0.00	6.58	1.97	0.66	3.29	1.97	0.00	0.00
17	51.52	18.18	6.82	0.00	0.00	0.00	10.61	0.00	0.76	9.85	0.76	0.00	0.00
19	32.29	28.13	6.25	0.00	0.00	0.00	16.67	0.00	2.08	11.46	2.08	0.00	0.00
21	35.00	16.25	2.50	0.00	0.00	0.00	33.75	3.75	3.75	2.50	0.00	0.00	0.00
23	27.63	15.79	7.89	0.00	0.00	0.00	26.32	2.63	3.95	7.89	1.32	1.32	0.00
Average (Major)	**37.55**	**19.80**	7.89	0.17	0.24	0.03	**19.50**	3.67	2.56	6.39	0.61	0.11	0.00
Minor Keys													
2	22.37	36.18	16.45	0.00	0.00	0.00	11.18	7.89	2.63	5.26	0.00	0.00	0.00
4	20.51	35.90	2.56	0.00	0.00	1.28	21.79	12.82	1.28	2.56	1.28	0.00	0.00
6	21.15	30.77	13.46	0.00	0.00	0.00	23.08	17.31	0.96	0.96	0.00	0.00	0.00
8	20.83	35.83	5.00	0.00	0.00	0.00	12.50	19.17	0.00	1.67	1.67	0.00	0.00
10	10.37	38.41	18.29	0.00	0.00	0.00	18.90	9.15	2.44	0.61	1.22	0.00	0.00
12	30.68	27.27	7.95	0.00	0.00	0.00	10.23	11.36	2.27	5.68	1.14	0.00	0.00
14	16.67	50.00	16.67	1.04	0.00	0.00	7.29	4.17	0.00	3.13	2.08	0.00	0.00
16	18.42	42.11	6.58	0.00	0.00	0.00	15.79	6.58	1.32	2.63	3.95	0.00	0.00
18	18.97	29.31	1.72	0.00	0.00	0.00	27.59	13.79	1.72	1.72	6.90	0.00	0.00
20	25.00	35.71	8.33	0.00	0.00	0.00	11.90	4.76	0.00	2.38	11.90	0.00	0.00
22	12.50	28.13	4.17	1.04	0.00	0.00	21.88	7.29	1.04	8.33	6.25	0.00	1.04
24	29.79	43.09	3.19	0.00	0.00	0.00	12.23	2.13	0.00	4.26	1.06	0.00	0.00
Average (Minor)	**20.61**	**36.06**	8.70	0.17	0.00	0.11	**16.20**	9.70	1.14	3.27	3.12	0.00	0.09
Average (All)	**29.08**	**27.93**	8.29	0.17	0.12	0.07	**17.85**	6.69	1.85	4.83	1.87	0.05	0.04

Table 5.5: Distributions of dissonant chord types in hand annotated data as a percentage of sequence length per prelude and totals.

Prelude	**9**	**maj9**	**min9**	**7b9**	**min7b9**	**dimb7b9**	**11**	**b11**	**b3b11**	**maj11**	**min11**	**13**	**maj13**
Major Keys													
1	0.00	0.00	0.00	0.00	0.00	0.00	**6.43**	0.00	0.00	0.00	0.00	0.00	0.00
3	0.00	0.00	1.92	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
5	0.00	0.00	0.71	0.71	0.00	0.00	1.43	0.00	0.00	0.00	0.00	0.00	0.00
7	0.36	0.00	1.79	0.36	0.00	0.00	2.5	0.36	0.00	0.00	0.00	0.00	0.00
9	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
11	1.39	0.00	0.00	1.39	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
13	0.00	0.00	0.00	0.83	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
15	1.32	0.66	0.00	1.32	0.00	0.00	0.00	0.00	0.00	1.97	0.00	0.00	0.00
17	0.00	0.00	0.00	0.00	3.03	0.00	0.76	0.00	0.00	0.00	0.00	0.00	0.00
19	0.00	0.00	0.00	0.00	0.00	0.00	1.04	0.00	0.00	0.00	0.00	0.00	0.00
21	0.00	0.00	0.00	5	0.00	0.00	1.25	0.00	0.00	0.00	0.00	0.00	0.00
23	0.00	0.00	0.00	3.95	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.32
Average (Major)	0.26	0.06	0.37	**1.13**	0.25	0.00	**1.12**	0.03	0.00	0.16	0.00	0.00	0.11
Minor Keys													
2	0.00	0.00	0.00	3.95	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
4	0.00	0.00	**5.13**	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
6	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
8	0.00	0.00	0.00	0.83	0.00	0.00	1.67	0.00	0.83	0.00	0.00	0.00	0.00
10	0.00	0.00	0.00	2.44	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
12	0.00	0.00	0.00	3.41	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
14	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
16	1.32	0.00	1.32	0.00	0.00	0.00	1.32	0.00	0.00	0.00	0.00	0.00	0.00
18	1.72	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
20	0.00	0.00	0.00	1.19	0.00	1.19	0.00	0.00	0.00	0.00	0.00	0.00	0.00
22	0.00	0.00	2.08	1.04	0.00	0.00	1.04	0.00	0.00	0.00	3.13	1.04	0.00
24	3.19	0.00	0.53	0.00	0.00	0.00	2.66	0.00	0.00	0.00	0.00	0.00	0.00
Average (Minor)	0.52	0.00	**0.76**	**1.07**	0.00	0.10	0.56	0.00	0.07	0.00	0.26	0.09	0.00
Average (All)	0.06	0.00	0.09	0.18	0.02	0.01	0.14	0.00	0.01	0.01	0.02	0.01	0.01

Table 5.6: Distribution of root pitch scale degrees in hand annotated chords relative to the main key and represented as semitones from the tonic.

Prelude	**0**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**
Major Keys												
1	28.57	0.00	14.29	2.86	0.00	17.14	2.86	31.43	0.00	5.71	0.00	3.57
3	27.88	1.92	14.42	0.00	5.77	5.77	0.96	24.04	2.88	8.65	2.88	4.81
5	22.14	1.43	15.71	1.43	8.57	9.29	7.14	18.57	1.43	12.86	0.71	2.14
7	21.43	1.07	16.43	0.36	10.00	11.79	1.79	17.50	1.43	13.57	2.50	5.71
9	22.92	0.00	17.71	2.08	10.42	10.42	1.04	22.92	1.04	10.42	3.13	3.13
11	16.67	2.78	16.67	1.39	6.94	12.50	0.00	13.89	4.17	16.67	1.39	6.94
13	20.83	0.83	13.33	0.00	9.17	2.50	5.00	18.33	2.50	15.00	1.67	10.83
15	19.74	3.29	18.42	0.00	4.61	7.24	4.61	25.00	3.95	7.89	0.66	5.92
17	24.24	0.00	15.15	0.00	5.30	6.82	3.03	25.76	0.00	12.12	0.00	9.85
19	18.75	0.00	16.67	0.00	10.42	7.29	4.17	19.79	0.00	14.58	0.00	8.33
21	27.50	1.25	16.25	0.00	5.00	3.75	2.50	28.75	0.00	12.50	0.00	6.25
23	19.74	0.00	17.11	0.00	7.89	9.21	1.32	22.37	3.95	15.79	0.00	2.63
Average (Major)	**22.53**	1.05	**16.01**	0.68	7.01	**8.64**	2.87	**22.36**	1.78	**12.15**	1.08	5.84
Minor Keys												
2	29.61	0.00	12.50	5.26	1.32	15.79	5.26	16.45	5.26	0.66	3.95	9.87
4	30.77	2.56	14.10	5.13	0.00	8.97	3.85	28.21	1.28	1.28	5.13	3.85
6	25.00	0.96	5.77	6.73	1.92	20.19	5.77	11.54	7.69	2.88	8.65	10.58
8	28.33	5.00	15.83	2.50	5.00	14.17	0.83	16.67	2.50	1.67	0.00	7.50
10	22.56	0.61	7.32	4.88	8.54	20.73	1.83	17.68	4.27	3.05	3.66	6.71
12	29.55	1.14	13.64	5.68	0.00	6.82	5.68	20.45	3.41	1.14	3.41	9.09
14	27.08	0.00	13.54	5.21	1.04	12.50	3.13	19.79	3.13	5.21	5.21	5.21
16	32.89	2.63	6.58	10.53	1.32	19.74	2.63	13.16	1.32	1.32	5.26	3.95
18	24.14	1.72	13.79	12.07	3.45	8.62	0.00	10.34	6.90	0.00	13.79	8.62
20	23.81	0.00	15.48	10.71	2.38	10.71	2.38	13.10	5.95	5.95	10.71	1.19
22	31.25	0.00	11.46	6.25	1.04	12.50	1.04	18.75	5.21	2.08	3.13	7.29
24	25.53	0.00	13.83	7.98	1.60	16.49	0.00	20.74	5.32	1.60	7.98	1.06
Average (Minor)	**27.54**	1.22	11.99	6.91	2.30	**13.94**	2.70	**17.24**	4.35	2.24	5.91	6.24
Average (All)	**25.04**	1.13	**14.00**	3.79	4.65	**11.29**	2.78	**19.80**	3.07	7.19	3.49	6.04

Chapter 6

Chords In Ornamental Music

In chapter 4 hidden Markov models designed to identify key and moments of key transition using chord sequences obtained from symbolic data and transcribed audio were described (Mearns et al. [2011]). This work, along with other statistical approaches to harmony (e.g.Rohrmeier and Cross [2008]), generates chord symbols for the Bach chorales, which can be derived with some accuracy due to the homophonic four voiced texture of the music. Alternative approaches work around the issue of inducing accurate chord data by basing their research on hand-annotated collections (e.g. Mauch et al. [2007], Anglade and Dixon [2008]), or online databases of manually created chord symbols (McVicar et al. [2011]).

To facilitate automatic harmonic analysis of a broad range of digital corpuses, a robust method of acquiring the underlying chords from ornamental polyphonic music is necessary. Accessing the chordal structures from rapidly moving instrumental textures is considerably more difficult than obtaining chord labels from homophonic works due to the presence of 'non-chord tones' - tones which are present in the musical surface but which do not form a part of the underlying harmony. For example, consider the excerpt from Bach's Prelude No. 7 shown in Figure 6.1. MIDI data for the first bar is shown in Table 6.1. From a systematic perspective, processing such information to deduce the implied harmony is complex on many levels. In Beat 1, the elaborated chord is an E♭ Major triad {E♭, G, B♭}. The non-chord tone in this group is the semiquaver A♭, situated in the treble clef, (alto voice), and passed over melodically in a progression which moves from the third (G) to the fifth (B♭) of the E♭ Major triad. The A♭ is metrically weak, although less so than the chord tones on either

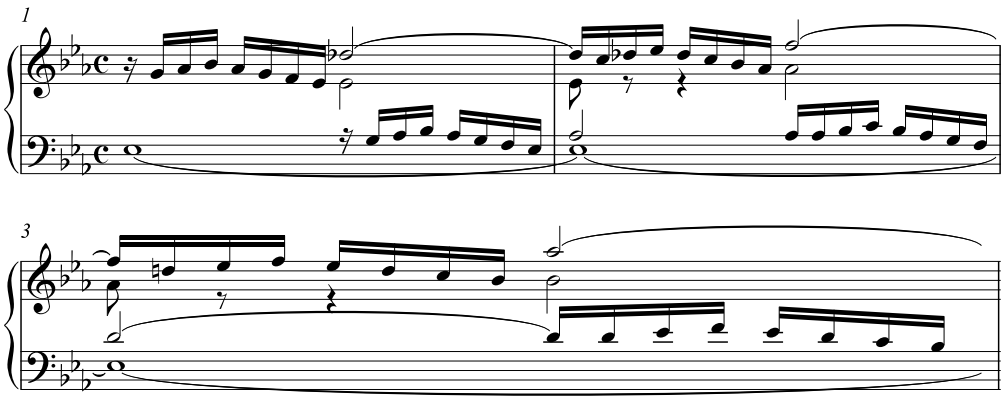


Figure 6.1: The opening of Prelude 7 of the Well Tempered Clavier, Book One, by J. S. Bach.

side of it. In Beat 2, the underlying chord is the same, and the non-chord notes are the A♭ and F. In both cases, to classify the implied chord the non-chord tones need to be discarded. In the second beat they occupy positions of greater metrical accentuation than the chord tones.

In Beat 3 the dissonant interval of a 7th is introduced in the form of an upper voice D♭ against an E♭ below. The dissonant interval is strongly emphasized in the musical surface by metrical position, (it commences on the strong beat on the second half of the bar), duration (two beats), and register (it takes place between the two highest pitches in the bar). The non-chord tone in this particular beat is the tenor voice A♭, in a figuration which echoes the opening semiquaver pattern of the alto.

From the previous example it can be seen that metrical position and duration are not unequivocal indicators of structurally important tones. Master composers rarely adhere consistently to singular rules of articulation; there are a huge variety of ways in which chords may be figurated in keyboard music. The ability to elaborate chords with variety and spontaneity was an indispensable skill of any working musician. (Friederich Erhardt Niedt's 'Musicalische Handleitung', published approximately 1710, and written purposely to improve variation techniques, has been mentioned earlier in chapter 5). In The Well Tempered Clavier there are many instances where chord tones are not presented at strong metrical

Table 6.1: MIDI and standard representation of note pitches of the opening of Prelude 7 of the Well Tempered Clavier, Book One, by J. S. Bach.

Segment Index	**MIDI Pitch**	**Musical Notes**
Beat 1		
0	51	E♭
1	51, 67	E♭,G
2	51, 68	E♭,A♭
3	51, 70	E♭,B♭
Beat 2		
4	51, 68	E♭,A♭
5	51, 67	E♭,G
6	51, 65	E♭,F
7	51, 63	E♭,E♭
Beat 3		
8	51, 63, 73	E♭,E♭,D♭
9	51, 55, 63, 73	E♭,G,E♭,D♭
10	51, 56, 63, 73	E♭,A♭,E♭,D♭
11	51, 58, 63, 73	E♭,B♭,E♭,D♭
Beat 4		
12	51, 56, 63, 73	E♭,A♭,E♭,D♭
13	51, 55, 63, 73	E♭,G,E♭,D♭
14	51, 53, 63, 73	E♭,F,E♭,D♭
15	51, 51, 63, 73	E♭,E♭,E♭,D♭

positions, particularly at moments when the chord tones are subservient to the voice-leading. A further area of complexity mentioned in reference to the creation of the hand-annotated data, is the antithesis between the humanistic and interpretive nature of musicology which allows multiple possibilities, and the rigorous systematic procedures of computer science which often requires a single result. In addition, for a chord algorithm to have general applicability, it must be capable of evaluating the relative structural emphasis of notes within a group of notes, and using this information to classify a broad range of chords including extended dissonances.

Table 6.2: Types of Inessential Tones [Hindemith, 1942].

Shorthand	**Brief Description**
W	Changing Tone
D	Passing Tone
V	Suspension
U	Unprepared Suspension
N-	Neighbour tone left by leap
-N	Neighbour tone approached by leap
A	Anticipation
F1	Unaccented Free Tone
F2	Accented Free Tone

6.1 The Problem of Chord and Non-Chord Tone Classification

It is often doubtful whether such cases involve real returning tones or broken chord formations [Hindemith, 1942] page 165

Hindemith [1942] devotes a section of his book chapter entitled *Harmony* to discussing the identification of 'non-chord' or 'inessential' tones. Hindemith's summarisation of the nine categories of inessential tones (reproduced in Table 6.2), along with his description of the different tone types and the influencing factors of the contexts in which they may appear, may give the impression that there are discrete and directly implementable rules enabling the identification of such tone types in complex musical textures. In practice, the opposite is the case. There are no absolute rules by which to define non-chord tones, and the rules that are suggested in music theoretic texts tend to take the form of guidance for the music analyst. Such rules are therefore imprecise, contextual and inherently circular. Hindemith, for example, states that in order to decide whether a tone is 'harmonic' or 'non-harmonic', the analyst must ascertain, based on a range of factors such as texture, harmony and metrical positioning, whether the tone produces an 'independent chord' or conversely, whether the tone is extraneous to the harmony (Hindemith [1942], page 164).

The opening to the second of Bach's French Suites (written between

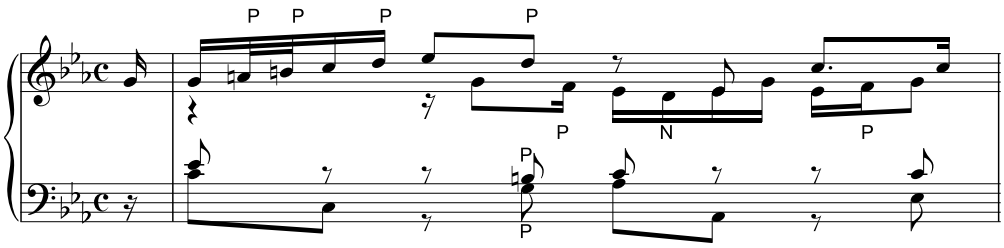


Figure 6.2: Suite II of Bach's French Suites, Opening Bar.

1722-1725), shown in Figure 6.2, provides a useful illustration of this issue of circularity. At the simplest level, a passing note, denoted by a 'P' in the diagram, is a note through which a melodic sequences passes from one note to another by step in a single direction, where a step is the interval of a semitone or tone. As may be seen from the illustration, more than one passing note may occur in succession in the linear sequence of tones (e.g. the adjacent A♮ and B♮ demisemiquavers in the upper voice of the opening figuration), and other types of inessential notes may occur simultaneously in other musical voices (e.g. the G, B♮, and D, on the off beat of the second crotchet beat of the bar). It can be seen from Figure 6.2, that a purely linear definition of a passing note is inadequate for the accurate identification of passing notes. Passing notes are only understood to be passing notes when they pass between *chord* tones, rather than *non-chord* tones. It is the harmonic implication of tone combinations that informs us that the adjacent A♮ and B♮ demisemiquavers are both passing notes, and inessential to the underlying harmony. The C to which they move could be defined as a passing note from a purely linear perspective, but we know that it is not, because it forms the root of the underlying C minor chord. The aforementioned G, B♮, and D group are a slight anomaly: at the quaver beat level these notes form the diatonic chord of G Major in the key of C Minor. In the example shown in Figure 6.2, all of the notes marked as passing notes occur at weaker metrical positions than the component chord tones. In contrast Figure 6.3 shows an excerpt from a sonata by Beethoven (written in 1796), featuring a chromatic decorative sequence, a device common amongst works of Beethoven's era (Classical

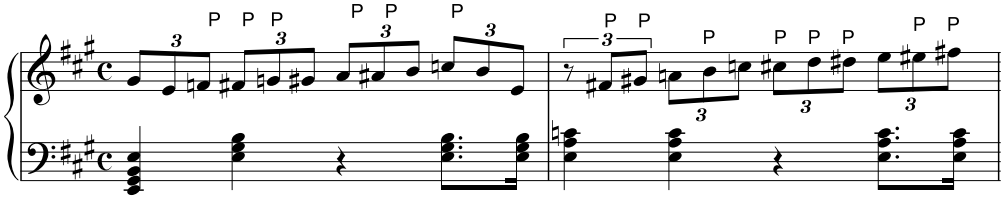


Figure 6.3: Beethoven's Sonata in A, Opus 2 No. 2, Rondo, Bars 89-90.

/ Early Romantic). In this example the chord tones in the sequence are picked out by matching their pitch to those contained in the sustained left hand chord, rather than by their metrical position or linear arrangement. The keyboard compositions of the Baroque historical period feature fewer such clear chordal signals - all of the musical voices are continually moving. Nonetheless, metrical emphasis of notes, particularly when combined with sustained duration, is anticipated to be an indicator of chord membership in particular situations (see section 6.2.5 below).

The lack of precision surrounding non-chord tone definition and identification in complex music presents a fascinating and difficult problem with respect to a computational implementation. The process involves the interpretation, translation and representation of fluid, inter-dependent, and inexact music theoretic constructs, into the rigorous form required for a computer program. The assumptions, translations and formal computational representations used in this research are now detailed in the following sections.

6.2 Guiding Principles from Music Theory for Tone Classification and Chord Recognition

The music theoretic principles of tone and tone feature recognition that commonly direct the judgements of music analysts and that have guided the computational implementation covered later in this chapter are described in this section. Hindemith's non-chord tone definitions are reduced for simplicity to two primary types of non-chord tone: passing notes, and neighbour notes (referred to as 'changing tones' by Hindemith), outlined

in subsections 6.2.1 and 6.2.2. Subsection 6.2.3 outlines the principle used to identify pedal notes. Subsection 6.2.4 discusses the principle of contour and its potential to express harmonic structure. Finally, the metrical theory of [Lerdahl and Jackendoff, 1983] which underpins the method of classifying the metrical emphasis of notes in this research, is detailed in subsection 6.2.5. These principles still demonstrate the issue of circularity mentioned earlier in this chapter. The computational approach to this problem and a more formal description of the classification methods are detailed in section 6.3.3.

6.2.1 Passing Notes

A tone is defined as a passing note if it satisfies all of the following conditions:

- Preceded melodically by step within a single musical voice
- Succeeded melodically by step within a single musical voice
- Is part of a melodic progression travelling in a single direction either up or down
- Does not form part of the underlying harmony

N.B. The definition implemented here does not recognize the non-chord tones denoted in Table 6.2 as *N*- and -*N*, i.e. passing notes approached or quitted by leap. Such notes are considered to be more difficult to spot and have been omitted for the sake of managing the level of complexity involved in the task.

6.2.2 Neighbour Notes

A tone is defined as a neighbour note if it demonstrates all of the following:

- Preceded melodically by step within a single musical voice
- Succeeded melodically within a single musical voice by the same note that preceded it

- Metrically weak
- Forms part of a group that does not cross a barline
- Is not adjacent to a passing note
- Does not form part of the underlying harmony

6.2.3 Pedal Notes

A tone is classed as a pedal note if it has:

- A total duration which is greater than the duration of a single bar

Pedal notes in the form of octave oscillations or repeated notes are not currently accounted for.

6.2.4 Contour

Music theorists, e.g. [Morris, 1998], have emphasised the concept of musical contour in twentieth century music, asserting that in the absence of tonal melody, contour is a prime structural feature in composition. It is possible that musical contour plays a similarly important structural role in earlier music, and that melodic peaks and troughs are used to accentuate notes. The theoretic principle of contour is therefore tested as a possible indicator of notes of structural prominence in the texture.

A tone is defined as occupying a contour peak (CP) or a contour trough (CT) if it is:

- Preceded melodically within a single musical voice by *two* steps or leaps either up (moving towards a CP) or down (moving towards a CT)
- Succeeded melodically within a single musical voice by *two* steps or leaps either up or down, but moving in the opposite direction to the preceding two steps / leaps.
- If the first note in a score is succeeded by *two* steps or leaps of the same direction it is marked as a CP or CT.

6.2.5 Metrical Structure

Lerdahl and Jackendoff [1983] explain that metrical emphasis is a function of the human perception of *beat*, rather than musical dynamic or patterns of articulation in the musical surface. (Please refer to chapter 2 for a more detailed explanation of the concepts of metre, beat and their relationship to simple and compound time signatures.) A *metrical hierarchy* is described, consisting of two or more levels of beats. The authors state that for a beat to be felt to be strong, it must 'also be a beat at the next larger level'. Figure 6.4 portrays an alternative representation of the metrical relations they describe, in which stronger beats (Lerdahl and Jackendoff's 'larger level') at higher positions in the diagram, and weaker beats are further down. The beats are represented diagrammatically using dots to symbolise equally spaced moments in time. Accented beats coincide with beats at the higher level above, a representation we consider to be more immediately intuitive than that of [Lerdahl and Jackendoff, 1983]. In $\frac{4}{4}$ meter, the beats on the first and third crotchet beats of the bar are felt to be stronger than the second and fourth beats because they are beats at the level above. Similarly the first beat is felt to be stronger than the third beat because it is a beat at the next level up. Figure 6.4 also shows the metrical accentuation levels of the simple triple time signature of $\frac{3}{4}$ and the compound duple time signature of $\frac{6}{8}$. In each case the strong beats are shown for each level in relation to the level below.

6.2.6 Chord Structure

The circularity of the harmony and non-chord tone problem is evident from the discussion in the preceding sections. In order to discover the underlying harmony in ornamental music, we wish to exclude, or at least reduce the emphasis of, non-chord tones from the chord recognition process. Moreover, to identify non-chord tones, we need to know that they are not part of the underlying harmony. Metrical information, surface emphasis, and a tone's linear positioning within a musical voice may assist in the identification of non-chord tones. In addition to this, the vertical intervallic relationship of a tone with other simultaneously occurring tones

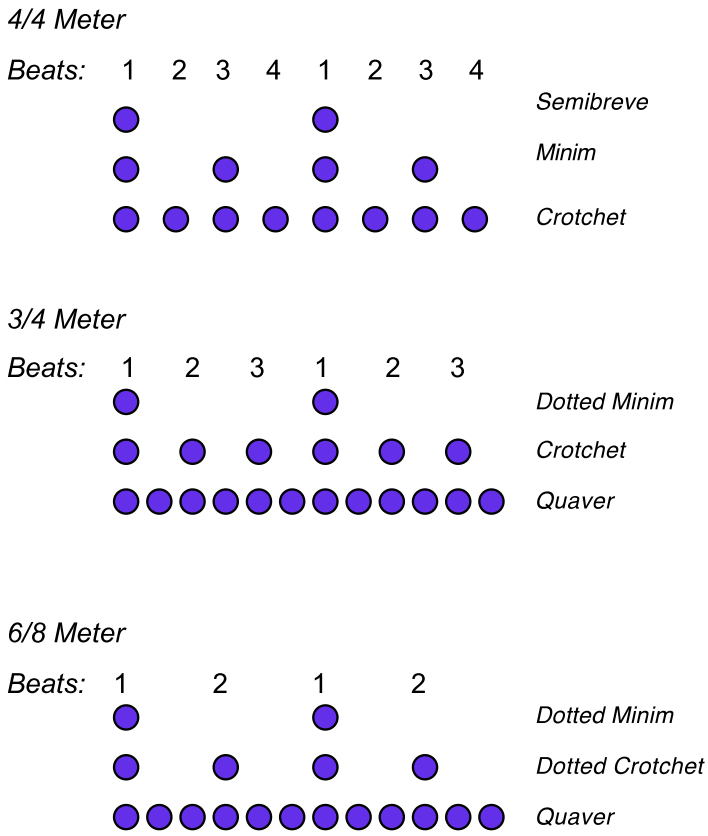


Figure 6.4: Expression of beat strength in a metrical hierarchy in which accented beats are also a beat at the level above.

must be taken into account, to ascertain whether a tone could be part of a chordal arrangement, even if it is not rendered prominent in the musical surface by dynamic, metrical, durational or melodic features.

The problem of chord definition has been discussed earlier in chapter 5. The chords we seek to discover are *tertian* arrangements; i.e. groups of tones that it is possible to organise into a series of successive *thirds*. The notes in a group are not necessarily presented so that they can easily be organised into a series of thirds moving upwards from the bass note. By re-arrangement however, chord notes may be formulated into a tertian group. The most common such arrangements are the major and minor triads (see the statistics of chord hand annotations in section Figure 5.10), followed by the 7th, and the diminished 7th. Due to the fact that chord

tones may be omitted from any chords, (a diatonic triad may be indicated by a single tone), but most particularly from extended chords such as the 7th, 9th, 11th and 13th, to meet the requirements of harmony notated for fewer conceptual voices, (for example four or three voice harmony), the component intervals of a chord are not always reducible to a series of successive thirds. They may also contain intervals representing an interval skip, or compound third (e.g. a sixth). Figure 2.6 gives an example of some commonly omitted tones in the arrangement of extended chords.

The principle used to identify a chordal arrangement from a vertical perspective is therefore to prefer groups of tones which can be organised into tertian or compound-tertian intervallic relationships from a submitted mixture of pitches. The full list of defined chords, interval profiles, and example notes are shown in Table 5.2.

6.3 Digital Score Processing and Note Feature Classification

This section of the thesis describes stages of digital score processing that form an integral part of automatic harmony analysis.

Section 6.3.1 describes the temporal segmentation of the music data. Section 6.3.2 describes the voicing method used to segment the data into linear musical streams. Section 6.3.3 describes the method of identifying passing notes in the voiced note data. Section 6.3.4 describes the identification of neighbour notes, contour notes, and pedal notes, in accordance with the principles listed in Section 6.2. The identified features and structures are then used to measure the notes' emphasis and importance.

6.3.1 Segmentation

The twenty-four preludes, symbolised in the form of MIDI data, are automatically segmented horizontally into linear musical voices as described earlier in this chapter (section 6.3.2). The voiced music data is divided into time segments which equate to a single musical beat in accordance

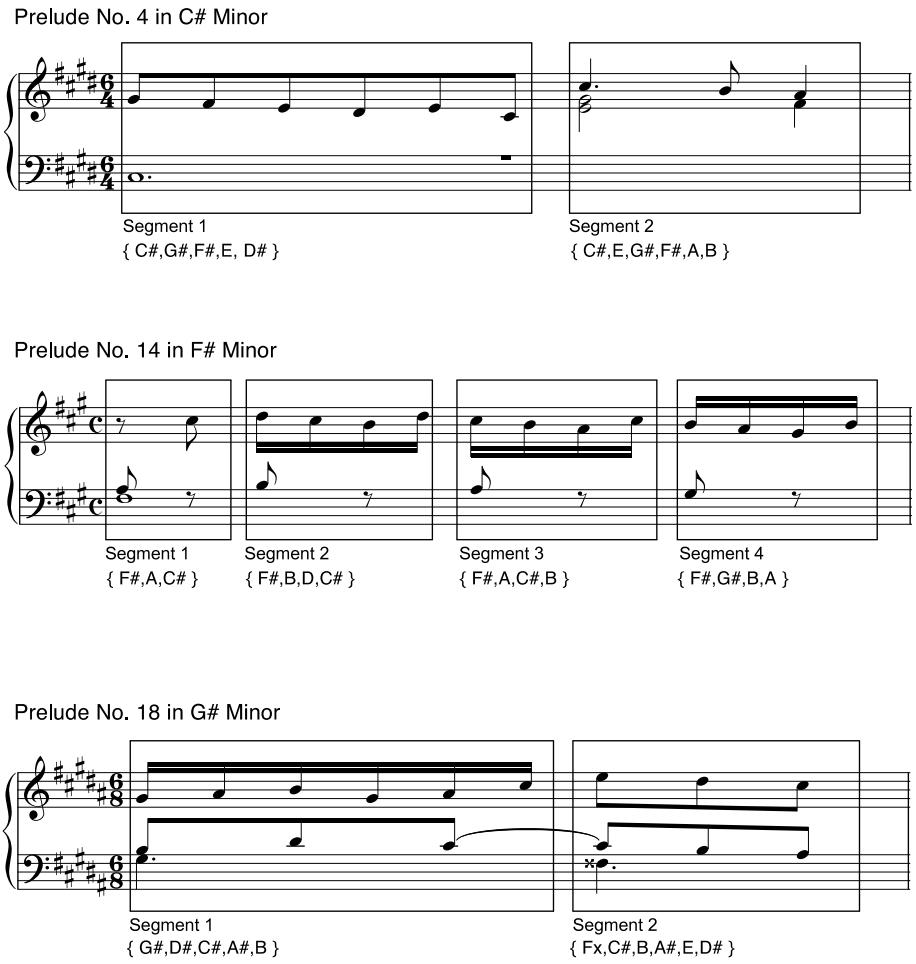


Figure 6.5: Segmentation and unique pitch sets for the opening bar of Preludes No. 4, No. 14 and No. 18.

with the time signature of the score. (Please refer to section 2.4 regarding metre and beat, and the discussion of harmonic rhythm in section 5.5.1.) A prelude in $\frac{4}{4}$ is therefore partitioned into four crotchet duration time segments per bar, a prelude with the signature $\frac{3}{4}$ is separated into 3 crotchet beat group segments per bar, and so on. The segmentation of three of the preludes is illustrated in Figure 6.5. Two of the preludes shown have compound time signatures; the prelude in C♯ Minor is in $\frac{6}{4}$, and the prelude in G♯ Minor is in $\frac{6}{8}$. The F♯ Minor prelude has a simple quadruple time signature. Boxes drawn around the groups of notes

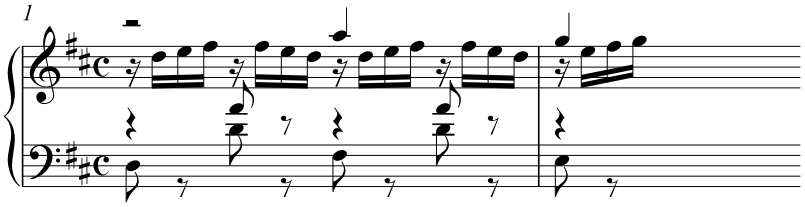


Figure 6.6: Prelude 5 in D Major, BWV 850, Bar 1, Riemann Edition.

in the score demarcate the beat segmentation; the unique set of pitches contained within each beat segment, (n.b. not showing pitch repetition), are listed within braces below the box.

Every individual musical note in a score is represented by a corresponding individual code object. Each beat group segment is allocated a list of object references for every note present during any part of that time segment. With reference to Figure 6.5, the first beat segment of the Prelude in C♯ Minor will contain seven note references: a dotted semibreve bass C♯, and six successive quavers in the treble clef: G♯, F♯, E, D♯, E and C♯. The second segment also contains references to seven note objects, including precisely the same bass C♯ code object.

6.3.2 Voicing Polyphonic Music

The concept of independent musical voicing in polyphonic music, the differences of opinion as to what constitutes a musical voice, and the intractability of the problem of automatically assigning accurate voicing to polyphonic music data have been discussed earlier in this thesis (see chapters 2 and 3 respectively). Considering the variations of voicing as annotated in different musical manuscripts (compare Figure 6.6 showing Riemann's scoring of Prelude 2, BWV 850, with the ABRSM edition in Figure 6.7), it can be argued that the problem is not fully solvable. Nonetheless, as may be ascertained from section 6.1, music data that has been voiced to a reasonable degree of accuracy when compared to an edition of a score is a prerequisite to the problem of non-chord tone identification.

The voicing method described in this chapter is loosely inspired by the

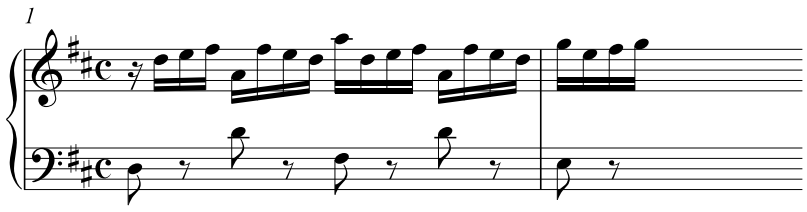


Figure 6.7: Prelude 5 in D Major, BWV 850, Bar 1, ABRSM Edition.

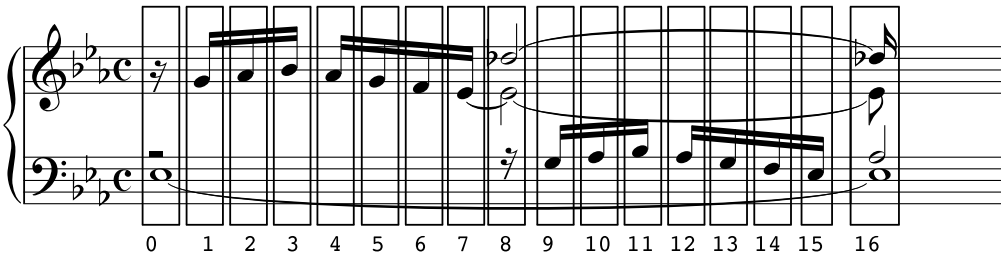


Figure 6.8: Vertical notegroup slices, Prelude 7 in E♭ Major, BWV 852, Bar 1.

contig mapping approach devised by Chew and Wu [2005] (please refer to Section 3.2).

The aim of the voicing method implemented here is to produce, as nearly as possible, a voiced version of an edition of a score, rather than to generate a voiced representation that is most closely aligned to the perception of a listener. The method initially segments the music data into a series of vertical note groups, (henceforth *VNG*), in accordance with each new note or rest onset. (N.B. 'empty' notegroups containing only rest values can be generated.) The vertical segmentation of the first bar of Prelude 7 in E♭ Major (BWV 852) is visualised in Figure 6.8. Rectangular boxes drawn around vertically coinciding notes indicate the vertical notegroups. Each note in a score is symbolised computationally by a single analogous note object. The bass E♭ with which Prelude 7 commences, sustained for two and a half bars, is represented by a single code object, and this object is referenced by the first forty eight *VNG's* each of which has a durational value of a semiquaver.

The method adopts the ideas of seeding maximal vertical note segments

Table 6.3: Summary of maximally voiced segments in the corpus: the total number of maximally voiced segments (MV), the maximum number of concurrent notes occurring in the work, and whether a maximally voiced segment features as the final chord.

BWV	Prelude	Key Key	No. of MV	Max No. of Voices	Final Chord
846	1	C Maj	1	5	Y
847	2	C Min	11	5	N
848	3	C♯ Maj	2	7	N
849	4	C♯ Min	1	6	N
850	5	D Maj	1	10	N
851	6	D Min	1	9	N
852	7	E♭ Maj	536	4	Y
853	8	E♭ Min	2	8	N
854	9	E Maj	17	4	Y
855	10	E Min	2	5	Y
856	11	F Maj	27	3	N
857	12	F Min	3	5	Y
858	13	F♯ Maj	347	2	Y
859	14	F♯ Min	15	4	Y
860	15	G Maj	1	4	Y
861	16	G Min	52	4	Y
862	17	A♭ Maj	1	5	N
863	18	G♯ Min	6	4	Y
864	19	A Maj	263	3	Y
865	20	A Min	5	5	Y
866	21	B♭ Maj	9	8	N
867	22	B♭ Min	1	9	N
868	23	B Maj	5	5	Y
869	24	B Min	2	4	Y

first and preferring voice connections with the closest pitch proximity, but departs in some significant aspects from Chew and Wu's method. Given that scores do in practice feature voice-crossing, voice-crossing is not penalised and it is possible for voice crossing to take place in this method. In addition, the maximum number of voices and voicing of large chords are interpreted differently. In contrast to Chew and Wu's method all notes in the score are voiced, irrespective of their contextual situation, be it homophonic or polyphonic, within a sparse texture or part of a large

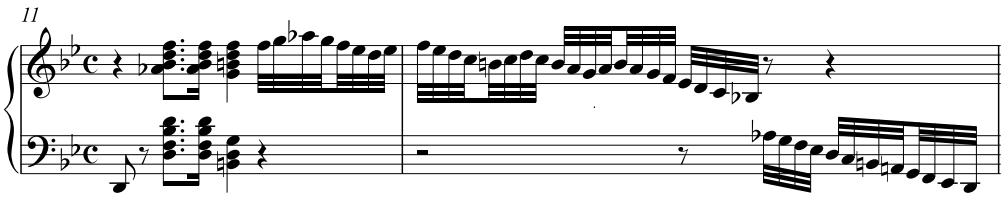


Figure 6.9: Mixed homophonic and elaborated texture, Prelude 21 in B♭ Major, BWV 866, Bars 11-12.

chord. Table 6.3 provides statistics about the vertical coincidence of notes in the test corpus. Out of the 24 preludes, 14 preludes feature a maximally voiced *VNG* as their final chord.

Figure 6.9, for example, shows the variegated texture of Prelude 21 (BWV 866), which alternates between dense homophonic chords and single lines of rapid demisemiquaver movement. How should these large chords be voiced? As a single voice of many notes? As the coincidence of many voices each only containing one note? The chords could be divided up into two or four voices, in accordance with predefined ranges representing soprano, alto, tenor and bass. But the demarcating such boundaries is problematic in itself. The chords and the rapid demisemiquaver runs evidence the importance of register on musical voicing, particularly with respect to scoring music. As can be understood by viewing Figure 6.9, by placing the most emphasis on voice connections of the shortest pitch distance, in the absence of concurrent pitches, a run of single successive notes such as the one commencing in the upper part of the treble clef in bar 11 in the figure, and concluding in the lower ranges of the bass clef at the end of bar 12, will be allocated the same voice irrespective of range, a result that could only be altered by explicitly accounting for registral position in the method.

The decision to allocate a voice value to every note in a score with no exclusions results in more voices than one might intuitively expect on looking at a score. Huron states a preference for a maximum of three voices based on perceptual experimental results [Huron, 2001], but the

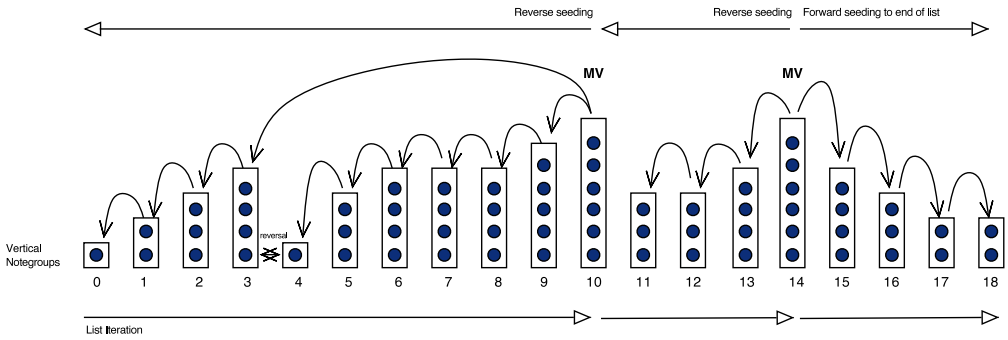


Figure 6.10: Voice Seeding Procedure.

extent to which composers considered the importance of the ability of listeners to accurately perceive formal contrapuntal processes can only be surmised. Composers, including J. S. Bach, recognise no such limitations on the maximum number of concurrent voices in a musical work, composing fugues of sometimes up to eight voices. The issue is returned to later on in this section, when a thresholding method to arrive at an optimum voice range is discussed.

An example of the overall voicing procedure is visually expressed in Figure 6.10. All maximally voiced segments (referred to as MV) are initially voiced in pitch height order in accordance with MIDI values (please see the MV annotated in the diagram). Maximal vertical slices containing duplicate pitches are not treated as maximal verticals but are left unseeded at this stage. Commencing with the first MV, the preceding *VNGs* are seeded using the method described below, by iterating back through the list in [X, Y] pairs, where X is seeded from Y, and where the first Y group is the MV and X is the notegroup preceding it in the list. The seeding process is indicated by arrows at the top of the diagram. The reverse iteration results in the majority of seedings taking place at boundaries where the number of voices is increasing, as per the results reported earlier Ishigaki et al. [2011]. (Prelude 1, for example, is voiced entirely in a single reverse sequence from the final 5 note chord.) In Figure 6.10 *VNG 10* is the first MV to be reached, and this group seeds the preceding group, group *9*. Group *9* voicing is then used to seed group *8*, and so on.

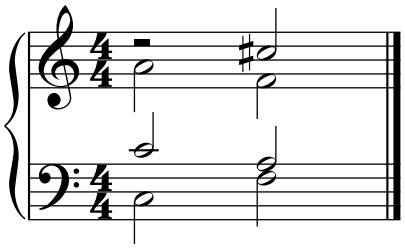


Figure 6.11: Musical example of X, Y notegroups, MIDI [48, 60, 69] and [53, 57, 65, 73].

The process continues in this way unless a 'reversal' situation is encountered, as demonstrated in the diagram between *VNG 4* and *VNG 3*. By *VNG 4*, the musical texture has reduced to a single voiced note which is inadequate for the task of informing about the voicing of all four notes of *VNG 3*. One possible solution to this is to allocate a voice to a single note in *VNG 3* according to the shortest distance from *VNG 4* and then order the remaining pitches in order of height and allocate voice numbers omitting the taken voice value. This approach does not take into account the registral positioning of *VNG 3* and in practice produces unusual levels of voice crossing and voicing values that do not tally with the score. An alternative approach, which improves the accuracy of results by taking into account the registral range of the nearest *MV*, is to seed notegroups positioned at reversals such as *VNG* 3 from the voice values of the nearest *MV*, in this case *VNG 11*. Once all the notes in a *VNG* have been voiced, the notegroup as a whole is marked as fully seeded. The seeding process continues either until the start of the list or until another *MV* is reached, at which point iteration continues from the next *MV*. Given that *MV* notegroups do not necessarily occur in the final chord in the music, when the final *MV* is reached, if it is not the last *VNG* in the score, the seeding process proceeds forwards in precisely the same way until end of the list.

To seed a pair of notegroups $[X, Y]$, all linear connections of pitches between the two groups are calculated, and a total voice leading score is

Table 6.4: All twenty-four possible combinations of X group [48, 60, 69] with seeding group Y [53, 57, 65, 73].

Index	Pair 1	Pair 2	Pair 3	Total Cost
1	48, 53	60, 57	69, 65	12
2	48, 53	60, 57	69, 73	12
3	48, 53	60, 65	69, 57	22
4	48, 53	60, 65	69, 73	14
5	48, 53	60, 73	69, 57	30
6	48, 53	60, 73	69, 65	22
7	48, 57	60, 53	69, 65	20
8	48, 57	60, 53	69, 73	20
9	48, 57	60, 65	69, 53	30
10	48, 57	60, 65	69, 73	18
11	48, 57	60, 73	69, 53	38
12	48, 57	60, 73	69, 65	26
13	48, 65	60, 53	69, 57	36
14	48, 65	60, 53	69, 73	28
15	48, 65	60, 57	69, 53	36
16	48, 65	60, 57	69, 73	24
17	48, 65	60, 73	69, 53	46
18	48, 65	60, 73	69, 57	42
19	48, 73	60, 53	69, 57	44
20	48, 73	60, 53	69, 65	36
21	48, 73	60, 57	69, 53	44
22	48, 73	60, 57	69, 65	32
23	48, 73	60, 65	69, 53	46
24	48, 73	60, 65	69, 57	42

arrived at by summing the absolute interval differences of each of the individual note pairs for that specific combination of connections. The score is used to find an overall shortest distance value of the two notegroups. For example, given two successive notegroups, where X consists of MIDI pitches [48, 60, 69] and Y has MIDI values [53, 57, 65, 73], 24 paired permutations are possible, as listed in Table 6.4. As can be seen from the musical scoring of the example in Figure 6.11, due to the position of the A♮ of the upper portion of the group, which is equally spaced between the F♮ and the C♯ of the Y notegroup, the first two sets of voice combinations result in precisely the same series of interval differences of 5, 3 and 4 semitones respectively, each with a total cost of 12. (Please refer to indexes 1

Table 6.5: Unison *VNG's* with a percentage of all *VNG's* in the preludes.

Prelude	**Total *VNG's***	***VNG's* with unisons**	**% unisons**
1	545	0	0.0
2	631	1	0.2
3	616	0	0.0
4	503	24	4.8
5	542	5	0.9
6	590	6	1.0
7	955	45	4.7
8	516	5	1.0
9	328	6	1.8
10	683	8	1.2
11	434	2	0.5
12	365	69	18.9
13	362	0	0.0
14	382	0	0.0
15	433	3	0.7
16	401	4	1.0
17	425	1	0.2
18	331	11	3.3
19	372	12	3.2
20	432	17	3.9
21	555	1	0.2
22	267	21	7.9
23	297	1	0.3
24	376	6	1.6

and 2 in the table.) In this event the algorithm selects the combination at the highest position in the array, in this case the voice connections, [48, 53], [60, 57], [69, 65]. The voice connections [48, 53], [60, 57], [69, 73] are equally plausible according to the scoring method of the algorithm.

A problem encountered during the seeding process concerns the seeding of notegroups containing unisons (see Table 6.5). Unison pitch occurrences usually feature notes of different durational values, however the seeding process takes into account only pitch values. Consequently, during the first pass through the *VNG* list, notegroups containing unisons where the duplicates are both unseeded are passed over. The note in the unison occurrence with a longer duration is seeded as the process continues to go

through antecedent groups, when the other unison value ceases to sound. To ensure that all remaining groups, including those containing unseeded unisons, are seeded, a second pass through the list searches for unseeded notes and assigns voice values in relation to the nearest *MV* followed by assigning the earliest available voice value remaining given the voice values already in use by the group, (i.e. if voices 1, 2, and 4 have been taken, even if the logical pitch distance voice for the unseeded unison is 2, the voice allocated to the note is 3 , as this voice value has not yet been allocated). (This latter in the event of an *MV* containing duplicates.) As an example, consider the excerpt shown in Figure 6.12 which shows bars 8 and 9 of Prelude 7, BWV 852. The final *VNG* of bar 8 contains three notes on B♭, with two voices sounding the B♭ below middle C, MIDI pitch 58. The resulting MIDI group is [46, 58, 58]. The seeding *MV* is the four note chord which strikes on the third beat of bar 9 with MIDI values [46, 57, 72, 75], voiced 1-4 from bass to soprano respectively. During the reverse seeding process, when the final slice of bar 8 containing the unison is reached, the seeding algorithm is unable to choose between the duplicate MIDI values of 58, and the group is therefore passed over without being seeded. On the next iteration the preceding *VNG* of bar 8 is reached (counting backwards) consisting of MIDI group [46, 58, 60]. This group references precisely the same sustained B♭ note (MIDI 58) as the unison group. It cannot be seeded from the unison group, therefore this group is seeded from the *MV* (MIDI [46, 57, 72, 75]), and correctly allocates voice 2, the same voice as the A♮ with which bar 9 commences, to the sustained B♭. The bass B♭ (MIDI 46) is already seeded as part of the *MV*, leaving only one note unseeded in the final *VNG* of bar 8. During the second pass of the list, the unseeded B♭ (MIDI 58) is seeded to voice 3 from the *MV* i.e. the same voice as the middle C demisemiquaver commencing the run in bar 9.

Voicing Evaluation Method and Results

To evaluate how the voicing method performs, the method is tested using a collection of MIDI recordings of the preludes in which musical voicing has

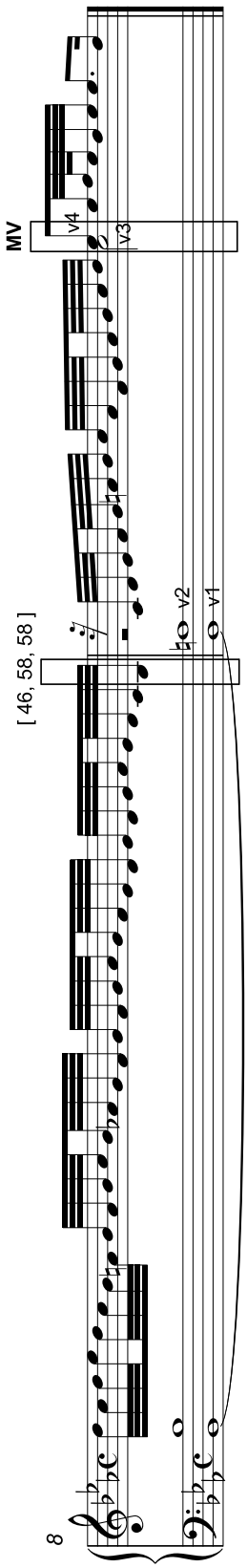


Figure 6.12: Unison pitch value example, Prelude 7 in E♭ Major, BWV 852, Bars 8-9.

Table 6.6: Number of tracks compared to MV in MIDI files.

Prelude	MIDI Tracks	Size of MV
1	3	5
2	2	5
3	2	7
4	4	6
5	4	10
6	4	9
7	4	4
8	3	8
9	4	4
10	4	5
11	2	3
12	4	5
13	2	2
14	3	2
15	2	4
16	4	4
17	3	5
18	4	4
19	3	3
20	3	5
21	3	8
22	4	9
23	4	5
24	3	4

been organised into MIDI tracks (aforementioned Phillip Kirlin dataset, section 5.8.2). The notes in the MIDI data are processed into musical voices using the method described above, and the voicing results are then compared to the corresponding track values in the original data. As can be seen from Table 6.6 however, there is a disparity between the maximum musical voicing value of each prelude and the number of channels allocated for voicing in the MIDI files. The primary reason for this is due to the previously described approach to large chords, in which each note in the chord is assigned a voice. Out of the 24 preludes, 6 MIDI files show an exact match between the number of MIDI channels and the number of voices. To calculate the precision of the voicing results of the algorithm

Table 6.7: Percentage of Matching Voice Connections to MIDI Ground Truth in the 24 Preludes.

Prelude	**% Matching Voice Connections**
1	99.7
2	99.4
3	99.8
4	97.6
5	98.8
6	99.2
7	98.9
8	96.1
9	99.4
10	99.0
11	100.0
12	97.5
13	100.0
14	98.7
15	98.6
16	99.1
17	98.7
18	99.4
19	98.6
20	97.1
21	90.5
22	97.5
23	99.3
24	99.6
% Average	98.4

the note to note linear connections are marked as either a match or a mismatch in comparison to the track allocations in the files. For example, if note x in the data is allocated voice 1, and note y is also allocated voice 1 by the algorithm, and note x is in MIDI track 4, and note y is also in MIDI track 4, then the voice connection would be counted as a match (1). In contrast, if note x in the data is allocated voice 1, and note y is also allocated voice 1 by the algorithm, and note x is in MIDI track 4 but note y is in MIDI track 2, then the connection is counted as a mismatch (0). The precision results for the twenty-four preludes are shown in Table 6.7.

Some of the preludes, notably prelude 21, demonstrate lower accuracies

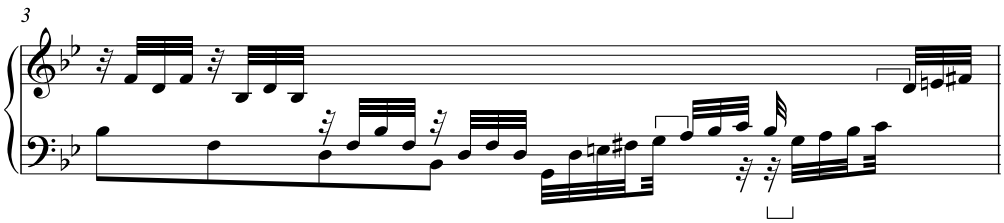


Figure 6.13: Differences between music notational voicing compared to pitch proximity voicing in Bar 3 of Prelude 21 in B♭ Major.

than the rest of the corpus. This is due to differences between the music notational voicing of the ground truth MIDI data and the pitch proximity voicing of the algorithm. Figure 6.13, which shows an excerpt from Prelude 21, gives an example of this contrasting approach - the three brackets in the figure show where the ground truth MIDI data changes voice but the voicing of the algorithm remains the same. The ground truth MIDI voicing is indicated by stem direction and accurately reflects the music notation in the score. The series of demisemiquavers occurring in the second half of beat 3 and situated on the bass staff, (A, B♭, middle C, B♭) changes voice in the ground truth data between the G and the A at the start of this 4 note sequence, and again at the end of it, following the B♭. In the algorithm's voicing these notes remain in the same voice; there is an absence of any other simultaneously sounding notes to inform the algorithm otherwise. From the perspective of Huron's perceptual rules, for example pitch proximity (see section 3.2), the voicing of the algorithm could be considered to be correct.

Thresholding to Obtain Optimal Voice Count

The voice texture of a musical work, as exemplified by the evaluation method, does not necessarily equate to the maximum number of concurrent notes or densest chord. To have the potential to automatically map the literal one note per voice values previously described across to a more credible range of voices, a method for deducing the optimum number of voices for a piece of music from the most dominant voice count is needed.

For example, a piece may be predominantly two voiced whilst incorporating a reasonably marked three voice section, in which case the optimal overall voice count would be three voices, despite three voices occupying a proportionally smaller role.

In this section we investigate the use of threshold values to select the optimal voicing. The experimentation does not address issues of compound voicing or the perception of musical voices, (for example, Prelude 1 in C Major could be considered to be a single harmonic arpeggiated stream rather than three independent voices/streams), but aims to discover how to identify the optimal musical voicing for a score. The method works by calculating the distribution of the number of voices per notegroup throughout each prelude. An upper and lower threshold percentage value are then used to select the optimum voice count. If the most common voice proportion is greater than the upper threshold percentage value, then this is chosen as the voice count for the piece, but if the second most common voice proportion is greater than the lower threshold and the second most common voice count is higher than the most common voice count, then this voice value is chosen. Maximal voicings of less presence than the lower threshold percentage value are discounted from the voicing value, (i.e. the presence of large chords dispersed in the musical texture does not affect the optimal voice count unless they occupy a significant section of the music). Table 6.8 shows the optimal musical voicing which results from using an upper threshold of 75% and the lower threshold of 15%. The results agree with the MIDI data in 15 out of 24 cases. For example consider Prelude 5, predominantly a light two voiced texture throughout, but featuring two very dense chords at the beginning of bars 33 and 34 (i.e. in the lead up to the final bar 35). The thresholding method identifies the two voiced texture of the prelude, despite the MV value of ten simultaneous notes being used in one of these chords.

6.3.3 Passing Notes

Each musical voice is initially processed separately, and notes in linear stepwise passing formations are labelled to indicate that the note may be

Table 6.8: Results of Threshold Method to Select Optimal Musical Voice Count Per Prelude.

Prelude	**Distributions of Voices (Number of voices, % of *VNG*)**	***MV***	**Optimum Voicing**	**MIDI Tracks**
1	(5, 0.73), (3, 99.27)	5	3	3
2	(5, 1.32), (3, 6.58), (1, 13.16), (2, 78.95)	5	2	2
3	(6, 0.32), (7, 0.65), (1, 5.48), (2, 93.55)	7	2	2
4	(6, 0.45), (5, 9.42), (2, 15.7), (4, 25.56), (3, 48.88)	6	4	4
5	(10, 0.73), (8, 1.46), (1, 2.19), (6, 2.19), (4, 2.92), (3, 8.76), (2, 81.75)	10	2	4
6	(6, 0.97), (7, 0.97), (8, 0.97), (9, 0.97), (1, 5.83), (3, 14.56), (2, 75.73)	9	2	4
7	(1, 0.36), (2, 7.22), (3, 28.16), (4, 64.26)	4	4	4
8	(7, 1.69), (8, 1.69), (6, 2.54), (1, 5.93), (2, 9.32), (3, 10.17), (5, 22.88), (4, 45.76)	8	5	3
9	(1, 0.72), (4, 6.14), (2, 20.94), (3, 72.2)	4	3	4
10	(5, 1.23), (1, 1.84), (3, 9.82), (4, 31.9), (2, 55.21)	5	4	4
11	(1, 2.31), (3, 6.94), (2, 90.74)	3	2	2
12	(5, 2.35), (3, 24.71), (4, 72.94)	5	4	4
13	(1, 4.19), (2, 95.81)	2	2	2
14	(4, 8.42), (3, 22.11), (2, 69.47)	4	3	3
15	(4, 0.23), (1, 11.09), (2, 88.68)	4	2	2
16	(2, 9.21), (4, 18.42), (3, 72.37)	4	4	4
17	(5, 0.77), (1, 5.38), (4, 8.46), (3, 13.08), (2, 72.31)	5	2	3
18	(4, 2.37), (2, 14.2), (3, 83.43)	4	3	4
19	(2, 14.74), (3, 85.26)	3	3	3
20	(5, 1.64), (1, 4.51), (4, 12.3), (3, 20.08), (2, 61.48)	5	3	3
21	(3, 1.25), (6, 1.25), (7, 6.25), (8, 6.25), (1, 41.25), (2, 43.75)	8	2	3
22	(3, 1.05), (7, 1.05), (9, 1.05), (6, 2.11), (5, 27.37), (4, 67.37)	9	5	4
23	(2, 1.33), (5, 2.67), (4, 12.0), (3, 84.0)	5	3	4
24	(4, 1.08), (2, 3.23), (3, 95.7)	4	3	3

a possible passing note, where M is the set of all possible passing notes. Given a possible passing note, $m \in M$, the labels *pred(m)* and *succ(m)* refer to the predecessor and successor respectively of each m within the voice. It is possible for a series of passing notes to be situated adjacent to one another. This initial stage adopts a purely linear perspective on passing note formations, and does not utilise information about inter-voice intervallic relations, duration or metrical position. A purely linear approach to passing note assessment does not provide a measure of the extent to which a possible passing note 'does not form a part of the underlying harmony', as detailed in the musical principles in section 6.2.1 and is inadequate for a corpus which features a great deal of consecutive stepwise movement. As discussed at the beginning of this chapter, the potential of a note to be a passing note must also be considered in the context of surrounding notes. Elements of M therefore have their classification refined using a scoring method that considers the chordal intervallic relationships of the note m, and the notes either side of it in the linear formation (*pred(m)* - m - *succ(m)*), in relation to the other pitches within the beat segment. The aim of the method is to verify whether a note marked as a potential passing note is harmonically essential or inessential in the context of the beat segment.

To ascertain this, the number of intervals indicating a potential chordal structure are counted for each of the three notes in the passing note formation. Specifically, the chord intervals (modulo 12) of unison, major or minor third, and perfect fifth, between each of the three notes in the passing note formation and all of the other notes present in the segment, are counted. Non-tertian intervals are not counted. Registral position is accounted for; if m is a D, and there is an A situated above this note, then the interval is classed as a perfect 5th, or 7 semitone interval. Conversely, if m has an A below it, the interval is classed as Perfect 4th, or 5 semitone interval, in relation to it. We recognise that allowing an interval in either direction is potentially counterintuitive and could result in a fifth being treated as a tonic, however the advantage is that it does not require identification of the tonic. We aim to improve upon the method and test the results systematically against a ground truth in the future (please

see chapter 7.) The result is a count of interval relationships for each of the three notes, represented as a 12 position vector where each position represents a semitone interval modulo 12. (Position 0 is a unison/octave, position 3 is a minor third, position 4 is a major third and position 7 is a perfect fifth.)

The interval score is used to improve the accuracy of the initial designation of m. In the event that *pred(m)* and *succ(m)* both have a greater quantity of chordal intervals in relation to the surrounding pitches than m, the scores are considered to evidence that the m note is harmonically inessential in context. Alternatively, if m has more triadic intervals than the notes either side of it, the evidence contests that m is harmonically essential. The score is used to determine the passing tone classification for each $m \in M$:

- If *pred(m)* and *succ(m)* both have higher scores than m, then m is removed from M and added to P the set of passing notes.
- If *pred(m)* and *succ(m)* both have lower scores than m, then m is removed from M.
- If either *pred(m)* or *succ(m)* has lower score whilst the other one has a higher score than m, then the classification of m remains, to indicate a possible passing note of less certainty.

For example, consider the first beat segment of Prelude 18, shown in Figure 6.5. The beat is a compound beat containing 10 notes in total: MIDI pitch values [68, 70, 71, 68, 70, 73, 59, 56, 63, 61], or musical notes [G♯, A♯, B, G♯, A♯, C♯, B, G♯, D♯, C♯]. The overall harmony is a G♯ minor triad. In the treble clef, the second semiquaver, A♯, is marked as a possible passing note (m). The pitches of the notes on either side of this A♯ are G♯ (*pred(m)*) and B (*succ(m)*). (The fifth semiquaver A♯ is not marked as a possible passing note because it is left by leap.) A count of interval relations between this A♯ (MIDI 70) and each one of the remaining notes in the beat ([68, 71, 68, 70, 73, 59, 56, 63, 61]) is computed, as shown in Table 6.9.

The resulting set of semitone intervals can be represented in an interval

Table 6.9: Interval relations of m, A♯, MIDI pitch 70, from the first beat segment of Prelude 18, (shown in Figure 6.5), to the other MIDI pitches in the segment: [68, 71, 68, 70, 73, 59, 63, 61, 56].

Note	**MIDI**	**Note**	**MIDI**	**Semitones %12**
A♯	70	G♯	68	2
A♯	70	B	71	1
A♯	70	G♯	68	2
A♯	70	A♯	70	0
A♯	70	C♯	73	3
A♯	70	B	59	11
A♯	70	D♯	63	7
A♯	70	C♯	61	9
A♯	70	G♯	56	(14) 2

Table 6.10: Vector representation of interval counts of G♯ - A♯ - B passing note formation notes in relation to surrounding pitches in the first beat segment of Prelude 18, Figure 6.5, along with the chordal interval score for each note.

Note (Semitones)→	**Pitch**	**MIDI**	**Interval Vector** [**0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11**]	**Score**
pred(m)	G♯	68	[2, 0, 2, 1, 0, 2, 0, 1, 0, 1, 0, 0]	4
m	A♯	70	[1, 1, 3, 1, 0, 0, 0, 1, 0, 1, 0, 1]	3
succ(m)	B	71	[1, 2, 1, 3, 0, 0, 0, 0, 1, 0, 1, 0]	4

vector as shown in Table 6.10. This table shows the interval count for all three of the notes in possible passing note formation (*pred(m)* - m - *succ(m)*) surrounding the first A♯ in this particular example. The number of chordal intervals present (unison, major or minor 3rd and perfect 5th) are totalled to produce a chord score for each note. The notes either side of the A♯ both result in higher scores, as shown in Table 6.10, consequently this note is moved to the set P, indicating that this note is a passing note and inessential to the surrounding harmony.

The C♯ in the alto voice on the third quaver beat of the first beat segment in Prelude 18 is also added to M. The notes either side of the alto C♯ are D♯ and due to the tie, B♮. This note is an interesting case

Table 6.11: Interval count and chord score of passing note formations in the opening bar of Prelude 18 (Figure 6.5.)

Note (Semitones)→	**Pitch**	**MIDI**	**Interval Vector** [**0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11**]	**Score**
pred(m)	D♯	63	[0, 0, 1, 0, 1, 2, 0, 3, 1, 0, 1, 0]	4
m	C♯	61	[1, 0, 2, 0, 0, 1, 0, 2, 0, 2, 1, 0]	3
succ(m)	B♮	59	[1, 0, 2, 1, 1, 0, 0, 0, 0, 2, 0, 2]	3

in point, because the B♮ is outside of the segment in question. In this case, the relationship intervals are computed with reference to the notes within the beat segment, as if the note in question was a part of that beat segment. The interval counts are shown in the second section of Table 6.11. In this instance the notes either side of m do not both result in higher scores, consequently this note remains in M. If the first A♯ in the segment is classified as a passing note it seems reasonable to suppose that other A♯'s in the segment should be classified similarly. At this stage of processing, the fifth semiquaver A♯ remains unclassified. To address this disparity, all of the notes in the segment are re-processed, and classified notes seed the classification of notes of duplicate pitch with either a lower or no classification within the segment, i.e. a note in P will seed a note of duplicate pitch outside of P, and a note in M will seed an unclassified note of duplicate pitch. Matching the pitch classification of duplicate pitch classes results in both C♯'s (treble and bass clef) being added to M and both A♯'s being added to P. The pitch duplication classification method means that some passing notes that are missed, perhaps because they are approached or left by leap, are successfully captured. In theory, by matching the classification of duplicate pitches in this way, the method should be able to deal with higher level harmonic abstractions.

Due to the lack of ground truth data, we are only able to evaluate the passing note method indirectly in section 6.4.8. There are situations when the scores of notes either side of a possible passing note are conclusively higher, very strongly indicating that the middle note is a passing note. Equally, there are situations when the scores of notes either side are higher

by only a small margin, indicating a less clear cut situation. There are also situations where either *pred(m)* or *succ(m)* produces a higher score but the other one does not. These situations are not currently addressed by the algorithm, but could be used to indicate varying levels of certainty of a passing note formation or passing notes approached or left by leap. The method would benefit from systematic research and evaluation, however creating the required ground truth data and performing comprehensive testing is beyond the scope of this PhD (please see future work in chapter 7).

6.3.4 Pedal Tone, Contour Tone and Neighbour Tone Classification

Pedal tones, contour tones, and neighbour tones are classified within a single musical voice from a linear perspective, and using rhythmic and durational information, based on the previously outlined rules. Neighbour tones must have durational values that are a fraction of a beat and are therefore captured within a single beat segment. Chordal arrangements are not taken into account in the identification of neighbour tones, i.e. the rule about forming a part of the underlying harmony is not implemented.

6.3.5 Implementing Measures of Metrical Strength

The aim of the implementation of metrical emphasis is to express the concept of stronger or weaker beats within a hierarchical metrical structure as outlined in Section 6.2.5. To betoken the differing proportions of metrical strength for a particular time signature, a series of metrical strength values are defined for quadruple/duple and triple time signatures. The strength values commence with a value of 1 for the strongest metrical position and decrease according to the positions' depth in the metrical hierarchy. For each metrical level, a value of 1/n is assigned, where n is the number of events in the bar at that level. A metrical position is assigned the value of the highest metrical level it occurs at. The different metrical levels and their numerical values in the metrical hierarchies can be viewed in Figure 6.14 for quadruple time signatures, and Figure 6.15 for triple. In practice

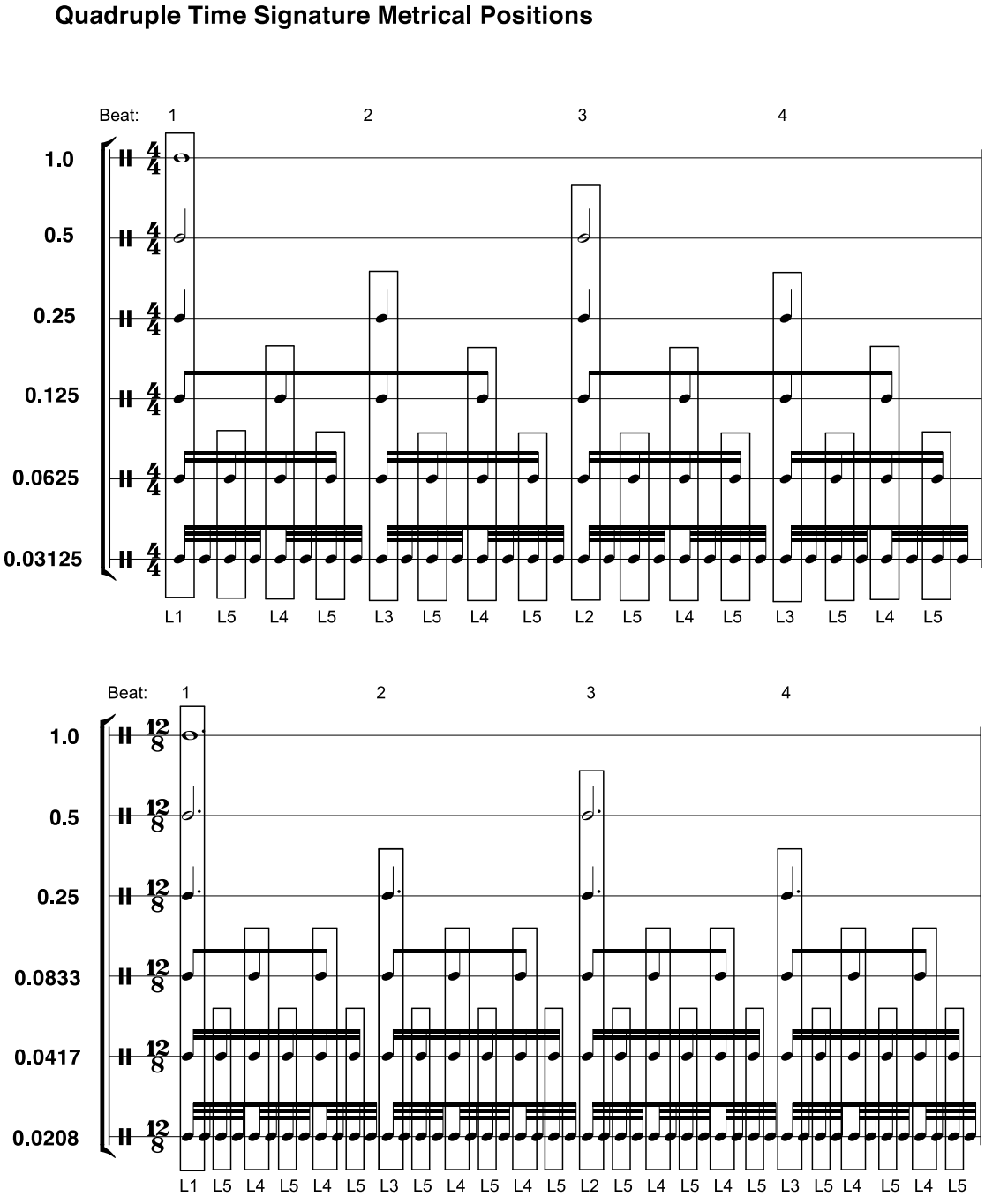


Figure 6.14: The hierarchy of metrical positions and values given for quadruple time signatures.

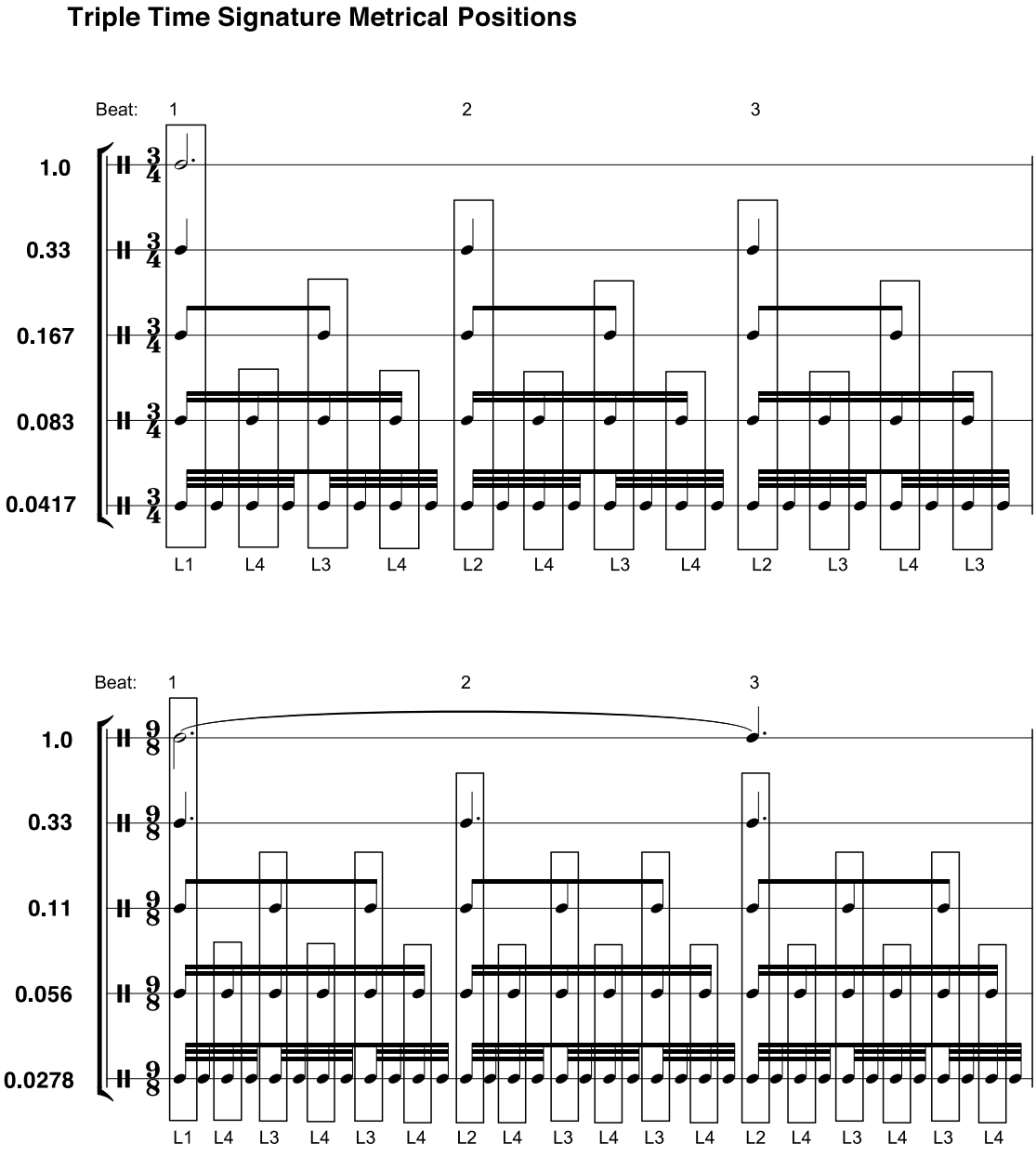


Figure 6.15: The hierarchy of metrical positions and values given for triple time signatures.

the smallest metrical level values are not used. For quadruple and duple time signatures the beat strength values are 1, 0.5, 0.25, 0.125, 0.0625 from strongest to weakest, and for triple time signatures the values are 1.0, 0.33, 0.16 and 0.08. For example, a note positioned on the first beat of the bar of a $\frac{4}{4}$ time signature is accorded a metrical strength value of 1. Beat 3, the beginning of the half bar, and the next strongest metrical level in the hierarchy, is given the next strength value in the series, 0.5. Beats 2 and 4 are positioned on the third strongest metrical level in the bar, and are allocated a value of 0.25.

6.4 Recognising Chord Tones using Note Features and Tertian Structure

In this work we hypothesise that harmonically important, or 'structural', notes are most likely to be strongly articulated in the musical surface, (i.e. emphasised by metrical position / duration / register / loudness etc.). Our hypothesis is supported by a large body of perceptual experimentation, extensively covered in [Deutsch, 1982], in which listeners are shown to mentally group notes into hierarchical representations based on metrical, registral, and tonal/temporal hierarchies heard in the music. Similar hypotheses have been made previously in computational approaches to musical structure, most notably *The Cognition of Basic Musical Structures*, in which Temperley states that structures such as phrase, counterpoint, meter and harmony are commonly accepted as 'musical facts', ('The Unanswered Question', page 1, [Temperley, 2001]). Temperley uses these 'facts' to support the musical intuitions on which his computational methods are based, further citing common practice music theory as evidencing the notion that musical 'salience' is linked to both theoretic and perceptual structure. A further hypothesis that we make in this work, is that information about chordal structure can also be extracted by computing the tertian intervallic arrangement potential of a group of notes. The importance of tertian structure in relation to musical harmony is supported by music psychology experimentation which suggests that humans

use arpeggiated or tertian patterns to formulate harmonic structure when listening to music ([Deutsch, 1982] and [Cook, 2006]).

The primary aim of this computational work is to capture the most plausible underlying or structural chord present during a particular beat segment, by simultaneously accounting for the structural emphasis of notes at the musical surface and the tertian intervallic relationships of those notes to one another. The overall approach therefore, is to combine heuristically derived measures of individual note emphasis, (arrived at by amalgamating evidence of note features, and metrical and durational proportions), with a calculation of optimum tertian note ordering and aggregation of the complete set of notes contained within a single segment.

This section outlines the approach used in more detail and formally describes the implementation details.

6.4.1 The Importance of a Note

A note whose surface articulation renders it prominent in the musical texture is construed as having increased likelihood of harmonic function within the context of a beat segment. For example, a note occurring on the first beat of the bar, sustained for a full beat, and situated at the peak or trough of a contour, is deduced to have a greater degree of harmonic function than an offbeat note of fractional duration marked as a passing note within the same beat. This concept, that the surface expression of a note can be related to the degree of its harmonic function, is referred to as the *importance* of a note.

To understand the effect of context on the role of a note, please see the opening bar of Prelude 4 in C♯ Minor, Figure 6.5. The bass C♯, although important in both the first and second segments, cannot be interpreted as having exactly the same degree of structural intensity in both segments, because the onset of the note occurs at the beginning of segment 1. The note is then sustained into segment 2. Although the bass C♯ is exactly the same note which happens to be present across two segments, its emphasis, and by implication its role in the context of the second segment, may

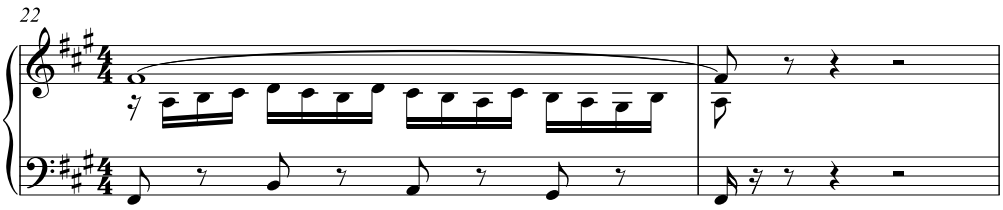


Figure 6.16: Prelude 14 in F♯ Minor, BWV 859, Bars 22-23.

conceivably be quite different in the second segment. The idea of note importance being relative to its situation is further exemplified in Figure 6.16, showing bars 22-23 of Prelude 14 in F♯ Minor, featuring a sustained F♯ semibreve in the upper voice. The F♯ initially reinforces the root note of the chord of F♯ minor, transforms into the fifth of the chord of B Minor in the second segment, becomes the root of the first inversion of F♯ minor in the third, and in the fourth and final segment of the bar transforms into a seventh over the G♯ and B to form a half-diminished seventh chord minus the fifth.

6.4.2 Measuring Note Importance using Duration, Metrical Position and Note Features

Notes commence with an importance value of 0 in the context of the segment. The scoring method computes a note importance value for each note based on durational and metrical emphasis and the previously delineated classifications of pedal, contour, neighbour and passing notes. The actual values used are heuristically derived and based on the musical intuitions of the author.

A measure of note emphasis is calculated first by summing the durational and metrical scores of the note. The durational value of a beat is the fraction of the beat that the note overlaps. For example, a quaver in the context of a minim beat is valued at 0.25, a quaver in the context of a dotted crotchet beat is allotted the durational value of 0.33, a quaver within the context of a crotchet beat is given the durational value of 0.5.

Notes sustained across the full duration of a beat are awarded the maximum durational value of 1, irrespective of the absolute duration of the note. The metrical strength of a note is assigned according to the onset time of the note and to the previously described metrical position values (see Section 6.3.5). The metrical emphasis of a note is only included if the onset of the note occurs within the current beat segment, otherwise the metrical emphasis value is 0. For example, a note in a $\frac{3}{4}$ time signature whose onset time occurs on the second crotchet beat is given the metrical value of 0.33.

Referring to the aforementioned bass C♯ in Prelude 4 in C♯ Minor, Figure 6.5, this particular note would score the maximum durational value of 1 for both the first and second segments. The metrical emphasis value would be set to 1 for the first segment, because the onset time of the note coincides with the first beat of the bar. The note emphasis score generated by combining the metrical and durational score is 2 for this segment. In the second segment the metrical score is 0 because the onset time of the note occurred in the preceding segment, resulting in a note emphasis score of 1.

The measure of note importance is initialised to the note emphasis value. Values are then either added to or deducted from this value in accordance with feature classifications as summarised below. The initial value sets are listed in Table 6.12.

- If a note is classified as a pedal tone, and the onset time of the pedal tone is not within the beat segment, the note importance value is reduced by the pedal tone penalty.
- If a note is classified as a passing note in M or P the importance value receives the relevant passing note penalty.
- If a note is classified as a neighbour note, the importance value is given the neighbour note penalty.
- If a note is classified as a contour peak or trough, the importance value is summed with the contour value.

Table 6.12: Summary of Note Features and Initial Heuristic Values

Note Feature	**Value**
Contour	2
Pedal	-2
Neighbour Note	-2
Passing Note (Weak certainty $m \in M$)	-2
Passing Note (Strong certainty $p \in P$)	-4

- If the note has no feature classifications, its importance value remains at the initialised note emphasis value.

6.4.3 Computing All Possible Note Combinations Per Segment

We assume that in any beat segment, (excepting rest segments), some non-empty subset of the notes in that segment has a harmonic function. To discover the most structural combination of notes, a combination we refer to as the *best note combination* or *BNC*, we compute a score for every distinct note combination or subset from an input group, accounting for both the surface articulation of notes as derived from note features, and the tertian interval relationships of the notes in the subset. For each individual combination of notes, an overall chord score is realised by summing the note importance values for the group, computing a maximum tertian interval content score, and then combining the two scores to provide an overall result.

The principle of computing all possible combinations of an input group is most easily understood by viewing Table 6.13. The table shows all combinations of an input group of five notes, {C♯, G♯, F♯, E, D♯}, from the smallest combination, in this case set to a minimum of one note, to the maximum group of five. The result is 31 note combinations. (Note combinations may feature notes with duplicate pitches, but each individual note has its own importance measure and therefore must be included in the method). Computing combinations has the potential to become extremely data intensive; an input group of $n = 20$ for example, will result in $2^n - 1$, or 1048575 combinations. Our input groups generally have much fewer

Table 6.13: All possible subsets of a group of 5 notes {C♯, G♯, F♯, E, D♯}, from the smallest combination to largest, shown with pitch names and pitch class equivalents.

Index	**Pitch Name**	**Pitch Class**
1	C♯	1
2	G♯	8
3	F♯	6
4	E	4
5	D♯	3
6	C♯, G♯	1, 8
7	C♯, F♯	1, 6
8	C♯, E	1, 4
9	C♯, D♯	1, 3
10	G♯, F♯	8, 6
11	G♯, E	8, 4
12	G♯, D♯	8, 3
13	F♯, E	6, 4
14	F♯, D♯	6, 3
15	E, D♯	4, 3
16	C♯, G♯, F♯	1, 8, 6
17	C♯, G♯, E	1, 8, 4
18	C♯, G♯, D♯	1, 8, 3
19	C♯, F♯, E	1, 6, 4
20	C♯, F♯, D♯	1, 6, 3
21	C♯, E, D♯	1, 4, 3
22	G♯, F♯, E	8, 6, 4
23	G♯, F♯, D♯	8, 6, 3
24	G♯, E, D♯	8, 4, 3
25	F♯, E, D♯	6, 4, 3
26	C♯, G♯, F♯, E	1, 8, 6, 4
27	C♯, G♯, F♯, D♯	1, 8, 6, 3
28	C♯, G♯, E, D♯	1, 8, 4, 3
29	C♯, F♯, E, D♯	1, 6, 4, 3
30	G♯, F♯, E, D♯	8, 6, 4, 3
31	C♯, G♯, F♯, E, D♯	1, 8, 6, 4, 3

than twenty elements, consequently the use of a combinations algorithm is an expedient application of the power of computing to the problem at hand.

6.4.4 Scoring Note Combinations from Note Features

To illustrate the scoring of note combinations, the scoring of the first beat segment of Prelude 4 in C♯ Minor, Figure 6.5, is detailed. The segment contains a total of seven notes; {C♯, G♯, F♯, E, D♯, E, C♯}. The underlying harmony in this opening chordal elaboration is C♯ Minor ({C♯, E, G♯}), consequently the F♯ and D♯ need to be discounted as inessential notes.

The previously described algorithms correctly classify the F♯ as a passing note, and the D♯ as a neighbour note. During scoring, these notes will receive the associated penalty. The C♯ and G♯ will both receive the highest metrical position score (1) and the bass C♯ is given the maximum duration score (1). The E's are awarded positive metrical position and durational scores in contrast to the negative values assigned to the inessential notes due to their position in the metrical hierarchy. For each distinct note combination of the 127 possible combinations the note importance values are summed and the top scoring group or groups of notes and associated score values are saved. The algorithm scores the five notes, {C♯, G♯, E, E, C♯} as the notes most strongly articulated in the segment, (alternatively the *BNC* for this segment using only note features), whose notes also match the pitch class set of the underlying harmony.

6.4.5 Scoring Tertian Arrangements of Note Combinations

To discover the optimum tertian arrangement of each distinct subset of notes, the pitches of the notes are converted into pitch classes, and all possible orderings of the unique set of pitch classes are computed using a permutation algorithm. Duplicate pitch values are not submitted: the aim is to find out whether the unique pitch class set can be organised in such a way as to form a tertian interval stack of notes. For example, Table 6.14 shows all twenty-four possible pitch arrangements of note combination No. 26 listed in Table 6.13: {C♯, G♯, F♯, E}. (The largest note combination in Table 6.13, {C♯, G♯, F♯, E, D♯}, produces 120 permutations and therefore was not used for this example.)

Table 6.15 shows the list of possible pitch arrangements of the chord of C Major ({C, E, G} = {0, 4, 7}). The successive interval content of

Table 6.14: The twenty-four permutations of note combination No. 26 from Table 6.13: {C♯, G♯, F♯, E}

Index	Permutation
1	C♯, G♯, F♯, E
2	C♯, G♯, E, F♯
3	C♯, F♯, G♯, E
4	C♯, F♯, E, G♯
5	C♯, E, G♯, F♯
6	C♯, E, F♯, G♯
7	G♯, C♯, F♯, E
8	G♯, C♯, E, F♯
9	G♯, F♯, C♯, E
10	G♯, F♯, E, C♯
11	G♯, E, C♯, F♯
12	G♯, E, F♯, C♯
13	F♯, C♯, G♯, E
14	F♯, C♯, E, G♯
15	F♯, G♯, C♯, E
16	F♯, G♯, E, C♯
17	F♯, E, C♯, G♯
18	F♯, E, G♯, C♯
19	E, C♯, G♯, F♯
20	E, C♯, F♯, G♯
21	E, G♯, C♯, F♯
22	E, G♯, F♯, C♯
23	E, F♯, C♯, G♯
24	E, F♯, G♯, C♯

Table 6.15: The permutations and successive semitone interval content of a C Major chord represented using musical pitch and pitch classes.

Index	Pitch Group	Pitch Class Set	Successive Intervals
1	C, E, G	(0, 4, 7)	[4, 3]
2	C, G, E	(0, 7, 4)	[7, 9]
3	E, C, G	(4, 0, 7)	[8, 7]
4	E, G, C	(4, 7, 0)	[3, 5]
5	G, C, E	(7, 0, 4)	[5, 4]
6	G, E, C	(7, 4, 0)	[9, 8]

each pitch permutation is calculated as if the note pitches are placed in ascending order. For example, when calculating the interval content of the third permutation in the table ({E, C, G} = {4, 0, 7}), the first pitch class of 4 is normally higher than the second pitch class of 0, consequently, to calculate the interval, pitch class 0 is represented by 12 rather than 0 to arrange the pitches in ascending order, and the interval difference between 12 and 4 is calculated). The result is a successive interval array for the permutation. The first permutation in Table 6.15 demonstrates that the pitch class set can be arranged into successive thirds: this permutation consists of the successive intervals of a major third followed by a minor third. This particular group of pitches can therefore be arranged into a tertian chordal stack, with the first permutation in the list showing the optimum tertian ordering of the pitch content of the group.

To select the maximum tertian arrangement of a pitch class set an interval scoring method is used. The method iterates through each interval in the successive array of intervals, and awards a score value based on the interval and its position in the interval array for each individual permutation of pitch classes. Intervals of thirds occurring in the first or second position of the array are awarded the *Triad Intervals* score. Intervals in position 3 of the array are awarded the *7th score*, in position 4, the *9th*, and position 5 or 6 the *11th and 13th*.

To mitigate against a bias towards the selection of the best mathematical arrangements of pitches in terms of thirds, (there are instances in the corpus where every pitch in a large combination of pitches can be organised into a series of successive thirds), rather than the most musically plausible result, pitch permutations resulting in an interval array consisting solely of thirds, (i.e. containing only values of 3 or 4 semitones) and of a total length of 4 or less intervals, are awarded an additional *Tertian Intervals Only* value. Interval arrays consisting of thirds and the doubled value of a third, i.e. intervals of 6 or 7 semitones which may indicate tone omission from the chord presentation, are awarded the smaller additional value *Tertian Intervals and Double Third*. The interval of 8 semitones is not used, because chords containing this interval tend to be beyond the range of the most common diatonic chords, please see Hindemith's table

Table 6.16: Summary of Tertian Heuristic Score Values

Chord Feature	**Value**
Tertian Intervals Only	2
Tertian Intervals and Double Third	1
Triad Intervals	2
7th	0.5
9th	0.25
11th and 13th	0.125
Double Third	1

of chord groups [Hindemith, 1942]. Please refer to Table 6.16 to see the full list of parameters set and their values, and Table 6.17 for an example of the interval scoring of the set of permutations of a G7 chord.

6.4.6 Selecting the Top Scoring Combination of Notes

Each distinct combination of notes for a given segment generates three score values:

1. Note Importance Score
2. Tertian Arrangement Score
3. Combined Note Importance and Tertian Arrangement Score

The best note combinations (*BNCs*) for a given segment are selected on the basis of the highest note importance score, the highest tertian arrangement score, and the highest combined score which is obtained by adding the note importance score and the tertian arrangement score together.

In the event that there is more than one equal top scoring group of notes for any of the three scoring categories, the choices are reduced iteratively to a single chord choice in accordance with the preference rules listed below, in order of presentation. As soon as a single note combination is realised, the iterative process is concluded and no further preference rules are applied.

1. Prefer the combination which has the most notes in it, (i.e. capture

Table 6.17: The score for each permutation of a seventh chord on G using the initial value set listed in Table 6.16. The highest scoring permutation is highlighted in bold.

Index	Permutation	Intervals	Interval Set	Score
0	(7, 5, 2, 11)	(10, 9, 9)	(9, 10)	0
1	(7, 5, 11, 2)	(10, 6, 3)	(10, 3, 6)	1.0
2	(7, 2, 5, 11)	(7, 3, 6)	(3, 6, 7)	4.0
3	(7, 2, 11, 5)	(7, 9, 6)	(9, 6, 7)	1.0
4	(7, 11, 5, 2)	(4, 6, 9)	(9, 4, 6)	2.5
5	**(7, 11, 2, 5)**	**(4, 3, 3)**	**(3, 4)**	**6.5**
6	(5, 7, 2, 11)	(2, 7, 9)	(9, 2, 7)	0.5
7	(5, 7, 11, 2)	(2, 4, 3)	(2, 3, 4)	2.5
8	(5, 2, 7, 11)	(9, 5, 4)	(9, 4, 5)	0.5
9	(5, 2, 11, 7)	(9, 9, 8)	(8, 9)	0
10	(5, 11, 7, 2)	(6, 8, 7)	(8, 6, 7)	1.0
11	(5, 11, 2, 7)	(6, 3, 5)	(3, 5, 6)	2.5
12	(2, 7, 5, 11)	(5, 10, 6)	(10, 5, 6)	0.5
13	(2, 7, 11, 5)	(5, 4, 6)	(4, 5, 6)	2.5
14	(2, 5, 7, 11)	(3, 2, 4)	(2, 3, 4)	2.5
15	(2, 5, 11, 7)	(3, 6, 8)	(8, 3, 6)	2.5
16	(2, 11, 7, 5)	(9, 8, 10)	(8, 9, 10)	0
17	(2, 11, 5, 7)	(9, 6, 2)	(9, 2, 6)	0.5
18	(11, 7, 5, 2)	(8, 10, 9)	(8, 9, 10)	0
19	(11, 7, 2, 5)	(8, 7, 3)	(8, 3, 7)	1.0
20	(11, 5, 7, 2)	(6, 2, 7)	(2, 6, 7)	1.0
21	(11, 5, 2, 7)	(6, 9, 5)	(9, 5, 6)	0.5
22	(11, 2, 7, 5)	(3, 5, 10)	(10, 3, 5)	2
23	(11, 2, 5, 7)	(3, 3, 2)	(2, 3)	4

the maximum number of notes, the group of pitch classes [0,0,4,7] should beat [0,4,7]).

2. Prefer the combination containing notes whose onset times match the onset time of the first beat of the bar.

3. Prefer the combination whose lowest note matches the lowest note of all of the notes in the combination (considered to be the bassnote).

4. Prefer the combination which contains more than one equal scoring permutation of the same pitches (frequently an indication of different arrangements of a diminished 7th chord).

5. In the event of multiple highest tertian arrangement or combined scores, prefer the combination with the highest note importance score.

Referring once again to the first beat segment of Prelude 4 in C♯ Minor, Figure 6.5, nine distinct note combinations generated from the notes in this segment correspond to the unique pitch class set {C♯, E, G♯}, (there are two C♯'s, one G♯, and two E's in the seven note input group). All of these combinations therefore result in precisely the same tertian permutation score, which in this case is also the highest scoring permutation. Using the first preference rule, the largest note combination of the tied-permutation results is selected. This is the five note group matching the pitch class set {C♯, E, G♯}; no further preference rule processing is necessary.

6.4.7 Final Output Lists of Best Note Combinations

For every beat segment in an input piece of music, a sequence of *best note combinations* or *BNC's* is generated in pitch class set format, based on the note importance score, the tertian arrangement score, and the combined score. The result is three individual lists of pitch class set representations of the most structural note combinations (*BNCs*), as calculated by the algorithm, for each of the three measures, for every prelude in the test set. The combined score list is expected to produce the best overall list of note combinations out of the three.

6.4.8 Evaluation

The method was evaluated by comparing the output sequences of *BNC's* to the hand-annotated harmony data. The hand annotations are converted to pitch class set format to enable like for like comparison. Notes such as pedal notes, listed in the hand annotations, are not added to the pitch class set of the ground truth chord for the segment. (Please refer to Section 5.6 about the annotation syntax.) For example, if the hand-annotated chord is C Major over a B♭ pedal note, the ground truth pitch class set is {0, 4, 7}, and the B♭ ({10}), is not represented.

Table 6.18: Tuned Values for Note Features

Note Feature	**Value**
Contour	0
Pedal	-1
Neighbour Note	-1
Passing Note (Weak certainty $m \in M$)	-1
Passing Note (Strong certainty $p \in P$)	-2

Table 6.19: Tuned Values for Tertian Score

Chord Feature	**Value**
Tertian Intervals Only	1
Tertian Intervals and Double Third	0.5
Triad Intervals	1
7th	0.5
9th	0.25
11th and 13th	0.125
Double Third	0.5

The combinations method does not fill in missing pitches, therefore, rather than penalising the method for pitches that are not actually played, we derive from the hand-annotated chord data a new ground truth which excludes pitches missing within a beat segment. Taking the intersection of the hand-annotated pitch class sets and the notes actually performed in any one beat segment, a series of new ground truth sets are created and used for comparison with the *BNC* sequences resulting from the three different types of measures. For example, if the hand-annotated set is {0, 4, 7} but the pitches actually played in the segment are {0, 7, 1, 6} then the new ground truth set is {0, 7} because pitch {4} is missing. In this example, if the pitch class set selected by the method is {0, 7} the result is counted as a match. If the method produces a set of pitches which do not match precisely, for example {0, 7, 6}, or {0}, the result is a mismatch, and it is counted as an error. In cases where the hand-annotated data offers more than one chord possibility as equally valid, if the chord method produces a match to one of the options, it is counted as a match.

Optimising the parameters in this type of work is a complex problem because the characteristics of the solution space are not known. In addition, due to the computation time involved, a complete grid search is not possible. We make the assumption therefore that the individual features are independent of one another and will not change in an irregular fashion in relation to differences in other features. To achieve a locally optimal result, working from the initial values listed in Table 6.12 and Table 6.16, individual note importance and tertian arrangement parameters are systematically varied singly, and the impact on the results are observed.The value of each individual parameter is successively varied between 0 and the first value for which results deteriorate in step sizes of 0.25. Testing is continued beyond the point of initial deterioration in order to establish whether a reversal might occur, however in all cases deterioration continues. After an initial pass of all of the parameters, in which the value at the point at which deterioration occurs is chosen, the same type of optimisation is repeated several more times in order to further ascertain whether the optimal values chosen for each parameter produces the highest overall result during evaluation. This is considered to be a more efficient approach than a complete grid search. The experimentation is continued until the no further improvement in the evaluation results is observed. The note and tertian parameters determined from the tuning stage are listed in Tables 6.18 and 6.19.

The average accuracy results following parameter tuning are shown in Figure 6.17. The graph exhibits the general tendency for the tertian arrangement method to outperform combinations arrived at using note features. As anticipated, the combined score produces the best results overall for the preludes, with average scores of 60.8%, 57.2%, 48.0% for the combined sequence, tertian arrangement sequence and note importance sequence respectively. An interesting result revealed by the tuning stage is that positively weighting note contour has no impact on the final accuracy scores and can be left at 0.

Two of the three lowest overall accuracy results are achieved from Preludes 4 and 18. Both preludes have compound time signatures, $\frac{6}{4}$ and $\frac{6}{8}$ respectively. These are the only two preludes in the set with these

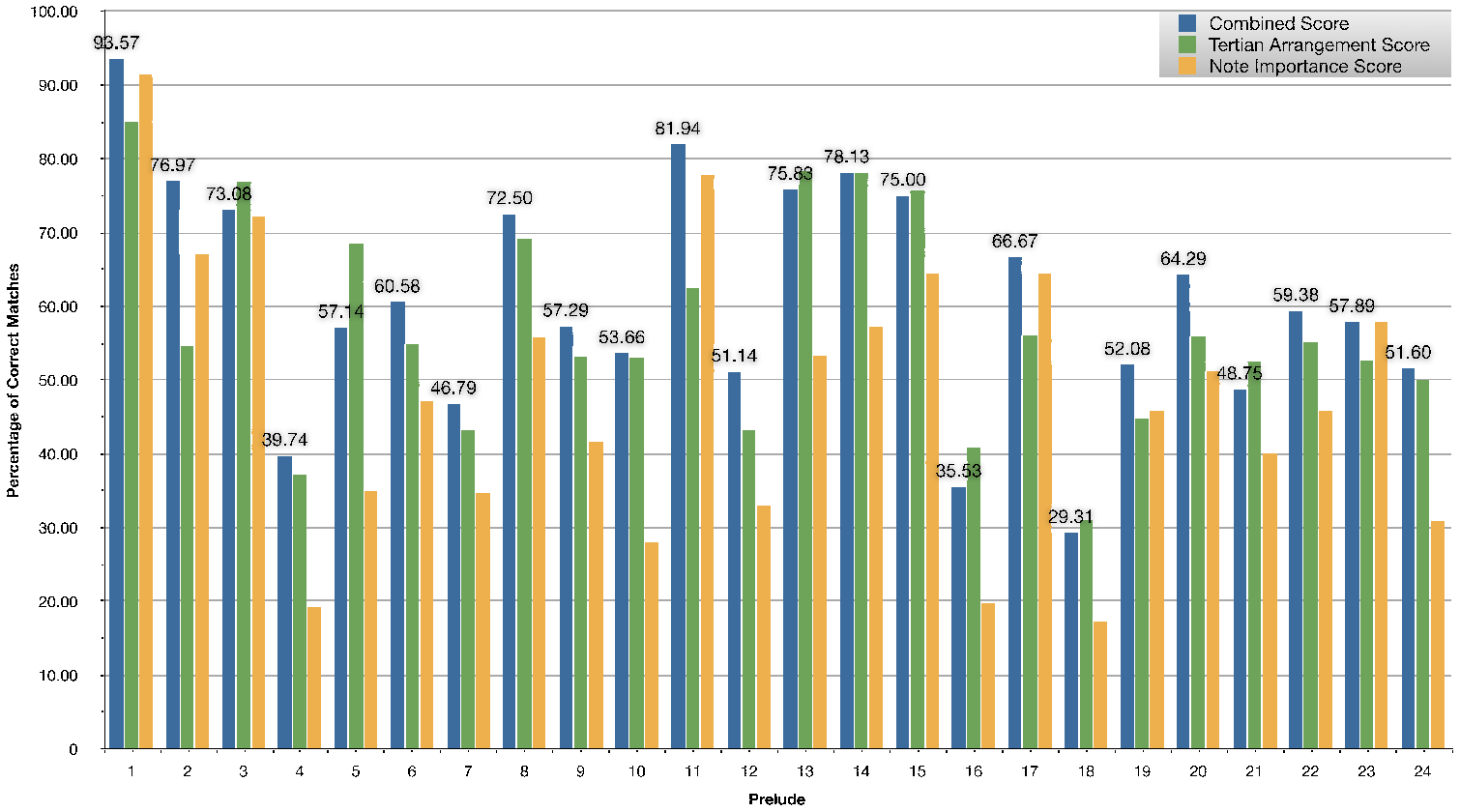


Figure 6.17: Average accuracy for the three sets of note combination sequences following parameter tuning. Values shown are for the combined sequence. Average overall values for the sequences are 60.78% (combined) 57.18% (tertian) and 47.96% (note importance).

particular time signatures, and in both cases the harmonic rhythm is frequently faster than the duration of a beat segment. (I.e. a single segment may feature a succession of two or more changes of harmony / possible chord labels.) The results demonstrate the weakness of the method in the abstraction of a single chord choice when several possible tertian arrangements of pitches of potentially equal validity are in contention. Rather, the method attempts to account for as many pitches as possible within a single chord label. The note importance weights, including metrical strength and passing note classification, are inadequate for the resolution of complex situations of multiple chord possibilities. Limited experimentation altering the metrical emphasis weights to more strongly favour notes occurring at the onset of a segment whilst negatively weighting notes which occur on weak metrical positions (e.g. occurring later on in the beat), marginally improves the chord abstraction accuracy of these two compound beat preludes, for example from 27.6% to 35.3% for the combined score for Prelude 18. A problem is that the alterations cause a not insignificant decrease in accuracy levels for the full set of preludes (approximately 8% lower average score). The issue is complex and requires more detailed and systematic experimentation, including potentially an extension to the method to automatically abstract harmonic rhythm (see future work).

It is anticipated that a change of segmentation to the lower value of a crotchet ($\frac{6}{4}$) and a quaver beat ($\frac{6}{8}$) will improve the accuracy of the results for these two preludes in the set. To ascertain the precise impact of compound segmentation, Prelude 18, consistently the lowest scoring prelude, is re-annotated at the quaver beat level, and the chord method is tested again. The consequence of changing the segmentation of the prelude to the smaller beat level proves to be considerable; much higher accuracy levels are achieved for the combined score, tertian arrangement score, note feature score: 70.7%, 67.2%, 66.1% respectively.

It is interesting to relate the note feature classifications, as described in Section 6.4.2, to the results of the method. The potential for note feature classifications relating to inessential note classifications (i.e. excepting metrical features) to influence the outcome of the method, varies

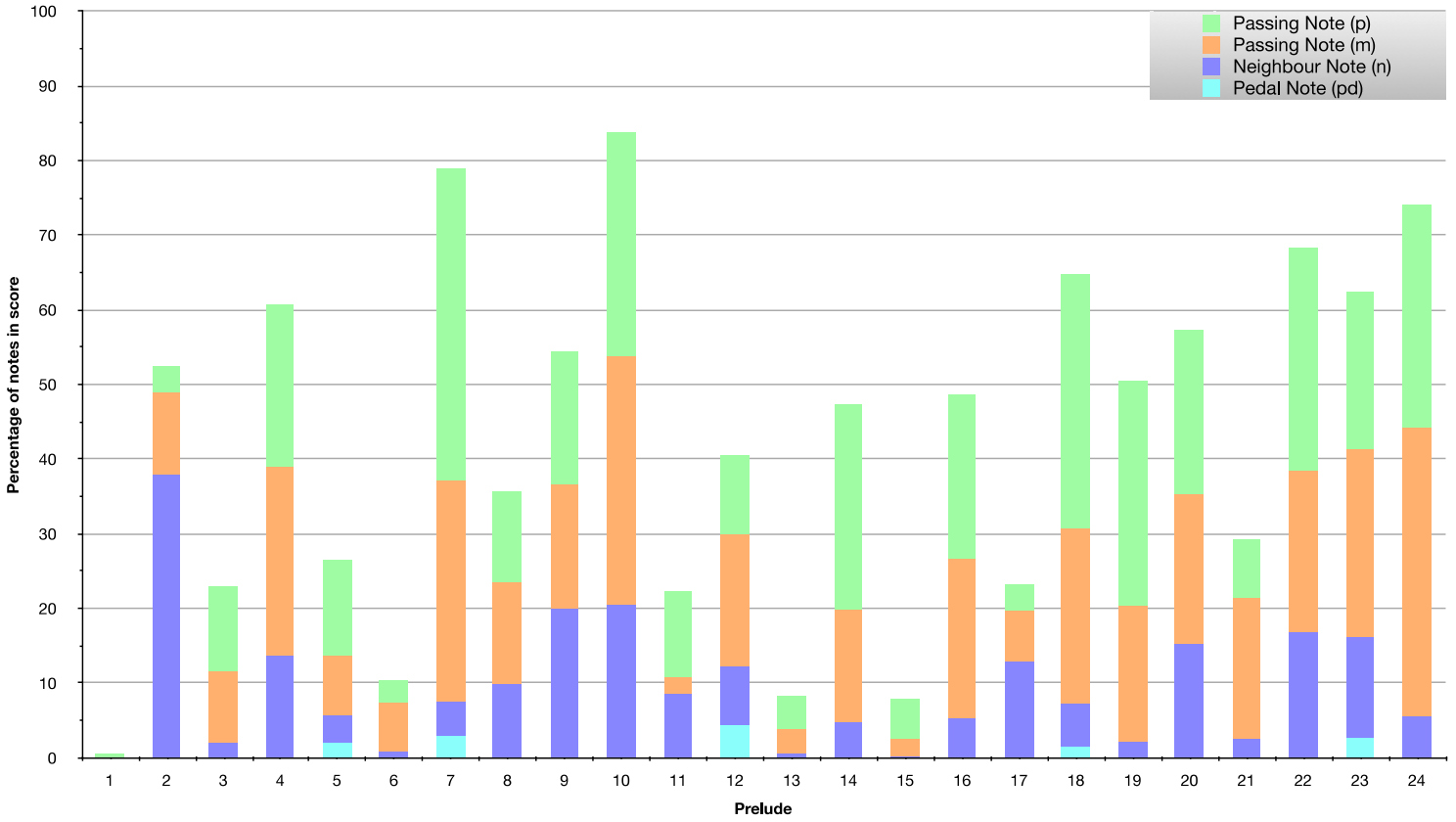


Figure 6.18: Distribution of inessential note feature classifications for the 24 preludes.

considerably depending upon the quantity of note feature classifications per prelude. Figure 6.18 details the inessential note feature classifications captured for each prelude by the previously described methods. Contour information has been omitted as it appears to have no effect on results unless weighted negatively (see above). Predictably, Prelude 1, with its simple arpeggiated figuration, produces the fewest inessential note classifications in the set, followed by Preludes 6, 13, 15, 17 and 21 in order of increasing classification percentage. The musical score editions of these preludes reveals that their melodic figuration is predominantly arpeggiated, much more so than the other preludes in the set. It could be deduced from this that these preludes contain the smallest quantity of non-chord tones and therefore it might be expected that these works yield the highest chord accuracy results. Relating the quantity and distribution of inessential note classifications, shown in Figure 6.18, to the chord accuracies shown in Figure 6.17, Prelude 1, probably the least harmonically ambiguous prelude of the set, consistently also gives the highest chord accuracy. Preludes 15, 13, 6, 11, 3 and 17, all have note feature classification quantities of less than 23%, and these preludes generate chord accuracy results ranging from 60% (prelude 6) to 82% (prelude 11), thus falling in the upper half of the accuracy range. To understand the correspondence between inessential note feature distribution and note combination accuracy, a scattergraph showing the relationship between inessential note classifications, (i.e. the percentage of the combination of neighbour notes, passing notes and pedal notes in the preludes) and accuracy, is shown in Figure 6.19, producing a Pearson coefficient of correlation of -0.6. The graph evidences an inverse relationship between inessential note features and the accuracy of method. Similarly Figure 6.21 plots the accuracy values in relation to the percentage of stepwise movement present, and shows a trend of greater accuracy in relation to fewer melodic steps for some of the corpus with a correlation coefficient of -0.62. Prelude 6, with its apparently minimal quantity of inessential tones, is an archetypical example of the challenges presented by master compositions to automatic computational analysis of music. Although the prelude is arpeggiated throughout,

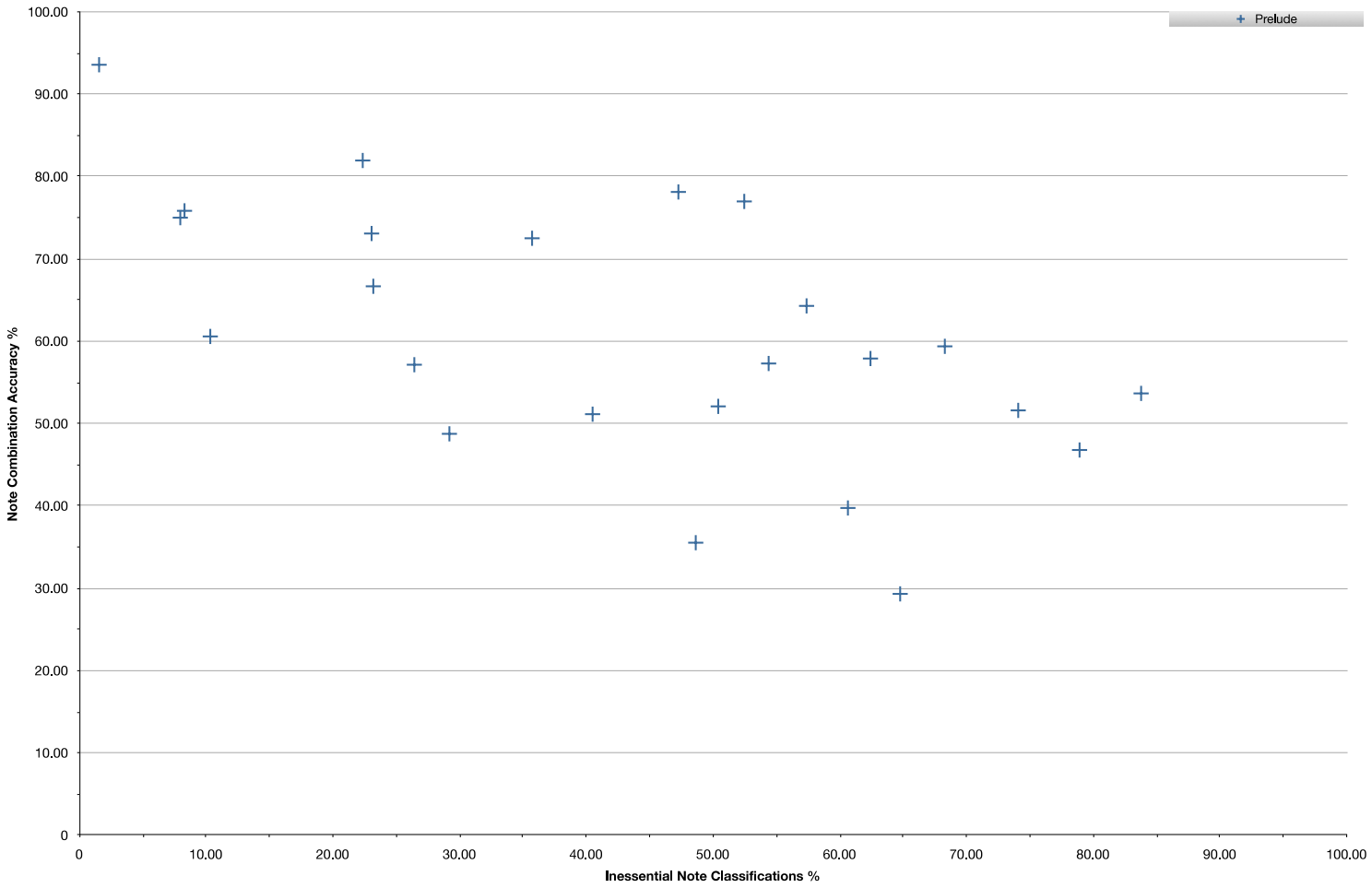


Figure 6.19: Average accuracy of combined sequence plotted in relation to the percentage of inessential note features (neighbour notes, passing notes, and pedal notes) calculated for the 24 preludes. The correlation coefficient is -0.6.

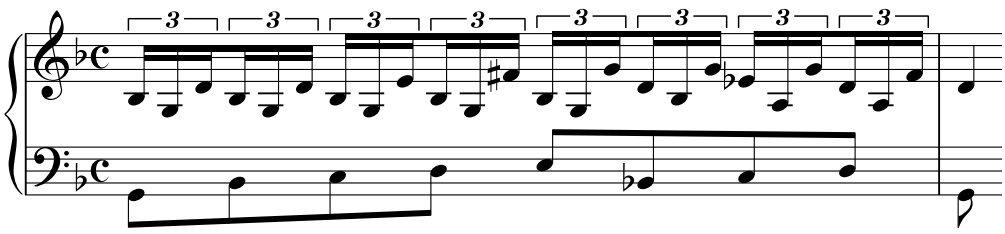


Figure 6.20: Prelude 6, Bar 7.

the figuration in the prelude simultaneously expresses both melodic movement and harmonic structure. Linear stepwise melodic movement, (often involving non-chord tones), is articulated at a higher level of abstraction. Voice-leading movement is promoted in the texture by positioning the melodic notes on the 3rd semiquaver of each tuplet in the treble clef. For example, see the ascending scale D, D, E, F♯, G in the treble clef of bar 7 of Prelude 6, in Figure 6.20. The software does not capture such higher level abstractions, consequently most non-chord tones are undetected in this prelude.

6.5 Labelling Note Combinations using Chord Dictionaries

6.5.1 Introduction and Baseline Evaluation

To arrive at a definitive sequence of chord labels per segment per prelude, the note combinations in the output sequences need to be awarded an unambiguous chord classification. The problem is by no means a trivial one. The note combinations themselves present two specific difficulties: firstly, the absence of critical chord components, either due to erroneous exclusion as a result of the previously described pre-processing, or, more frequently, due to the omission of such tones in the original music; and secondly, the presence of non-chord tones that the previously outlined methods have not successfully excluded. The test corpus, characterised throughout by complex ornamentation, features a multitude of instances of chords implied by fewer tones than the component tones of the chord,

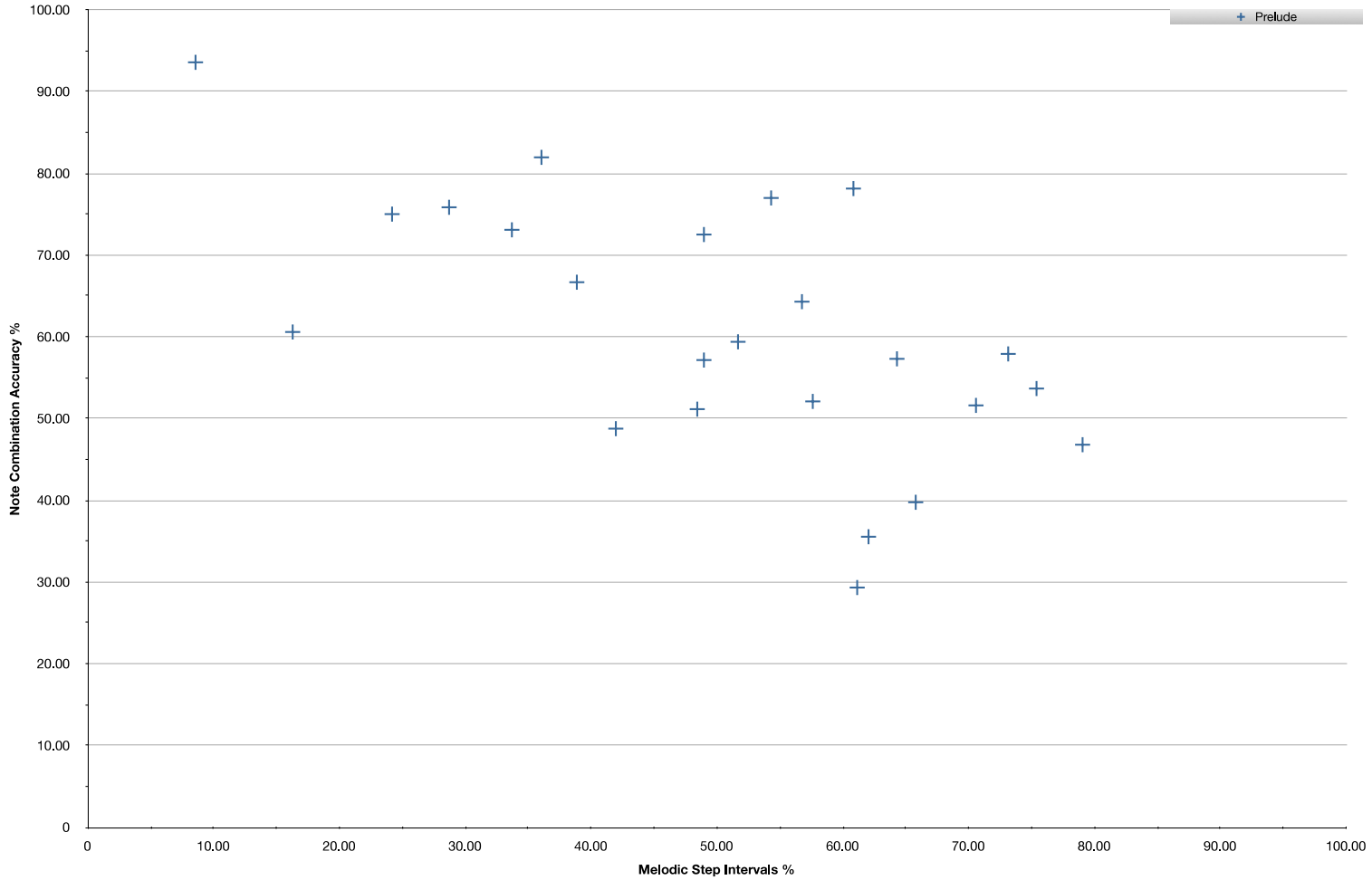


Figure 6.21: Average accuracy of combined sequence in relation to the percentage of melodic steps in the 24 preludes. The correlation coefficient is -0.62

(for example, a dyad or single tonic note implying the tonic triad), that are also decorated melodically with runs, trills, turns and so on.

To procure a distinct chord label per note combination, a chord dictionary is used, similar to that of Pardo and Birmingham [2002]. They describe a method (henceforth referred to as the *Harman* method), of classifying groups of pitch classes by scoring them against fully factored pitch class set chord templates. (For example, the major triad on C, {C, E, G}, is notated as pitch class set {0, 4, 7}, and the seventh on G, {G, B, D, F}, is notated as {7, 11, 2, 5}.) Their method counts the weight of notes in an input segment matching a template element (positive evidence), not matching a template element (negative evidence), and the number of template elements missing from the input segment (misses), in order to select the highest scoring template.

Pardo and Birmingham [2002] acknowledge the problem of the generation of multiple equally top scoring chord templates per input segment, and outline a set of preference rules to reduce multiples to a single chord choice. They also state that their method, expounded using a chord dictionary of triads and two types of seventh (dominant and diminished), is easily extensible to the representation of complex chords. In practice, this is not the case. Correctly classifying pitch class groups denoting complex chords against a dictionary containing extended chord templates is a peculiarly challenging problem. Consider, for example, the fully factored representation of a thirteenth chord, {G, B, D, F, A, C, E}, or pitch class set {7, 11, 2, 5, 9, 0, 4}, compared to the most common thirteenth arrangement in a four voiced musical texture, {G, B, F, E}, or {7, 11, 5, 4}. Sparse input pitch groups such as these produce multiple equal scoring chord options that may not even include the desired classification. Should any non-chord tones also be present, the possibility of accurately identifying the extended chord becomes even more remote. Furthermore, as evidenced below, each extension to the dictionary exacerbates both the difficulty of accurately matching complex chords and the production of multiple top-scoring templates generated by a single input group.

The corpus is tested against a prototype of *Harman* system. Four different chord dictionaries are used; the first is based on the major, minor,

diminished and augmented triads only, the second populates the dictionary with the same set of triads and all eight types of seventh chord, the third adds six types of ninths, and the final dictionary includes all of the previous chords plus 11ths and 13ths. Chord templates are generated on every degree of the scale using the interval profiles of the defined chord types. (For the list of chord types please see Table 5.2.) The prototype varies from Pardo and Birmingham [2002] in the implementation of their final preference rule, which selects a single option from a group of matching diminished 7th chords that vary in terms of inversion, based on the resolution of the chord; the rule is not adaptable to a complex texture with a moving bass part and is therefore omitted.

Figure 6.22 shows the number of times the prototype method results in more than one chord option per segment per prelude as a percentage of the total number of segments in the list and in accordance with the chord dictionary used. In some cases, the number of multiple equal scoring templates is large, producing more than five possibilities. The preference rules have not sufficiently reduced the chord choices to a single result. The average number of multiple choices per dictionary, shown in the final column of the graph, are 40.98, 39.63, 28.69, 34.85 (triads,, triads and 7ths, triads 7ths and 9ths, and all templates respectively), with a standard deviation of 14.2, 11.6, 11.3 and 9.1 in the same order. The ability of the *Harman* method to label segments in the test corpus was evaluated by comparing the pitch class sets generated by *Harman* and the hand-annotated data and counting the number of exact pitch class set matches. The evaluation is lenient; in the event that there is a large number of multiple chord options, should one of the options match the hand-annotated data, it is counted as a match. Ideally these multiples would be reduced to a single possible answer. The specific chord template dictionary employed by the automatic method limits the range of possible chord choices to a varying extent, sometimes to a smaller set of chords than is represented in the hand-annotated data. The evaluation method accounts for this discrepancy by iteratively reducing the size of the chord in the hand-annotated data until it matches the contents of the chord dictionary being used in any particular instance. For example, if the hand-annotated pitch class

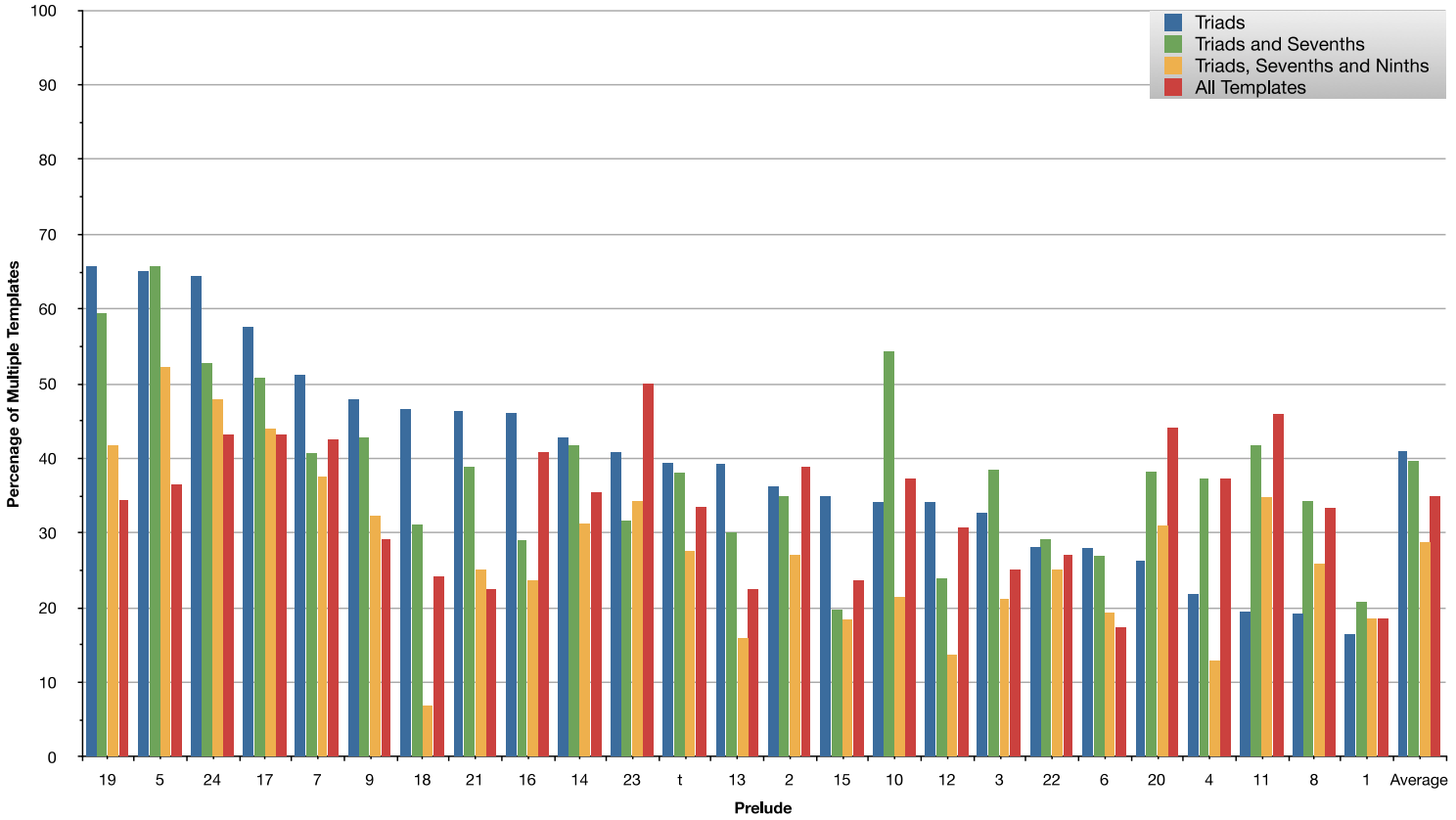


Figure 6.22: The percentage of segments using the *Harman* method [Pardo and Birmingham, 2002] resulting in multiple equal top scoring chord templates shown per chord dictionary.

Table 6.20: Chord accuracy results using the *Harman* method [Pardo and Birmingham, 2002] with four different chord dictionaries

Prelude	Triads	Triads and 7ths	Triads, 7ths and 9ths	All Templates
1	79.5	73.5	73.5	82.9
2	77.9	29.7	11.2	9.2
3	84.5	62.8	52.0	51.0
4	66.7	28.4	5.4	2.6
5	78.7	58.5	25.7	9.3
6	71.2	52.0	39.4	36.5
7	64.4	40.6	22.4	13.6
8	86.1	53.5	41.7	29.2
9	63.0	43.8	28.1	19.8
10	60.0	29.4	12.2	4.3
11	92.1	55.7	36.6	11.1
12	69.0	36.5	21.6	15.9
13	70.1	62.2	52.5	47.5
14	71.3	56.3	28.1	8.3
15	73.4	65.8	61.3	59.2
16	60.0	31.5	13.5	5.3
17	73.0	46.5	31.5	28.8
18	58.6	28.1	8.6	1.7
19	73.4	44.2	8.4	5.2
20	86.2	42.7	25.3	9.5
21	63.4	48.7	42.3	40.0
22	68.2	45.5	33.7	14.6
23	74.4	58.3	38.7	13.2
24	49.7	33.3	21.9	19.2
Average	**63.0**	**47.2**	**30.7**	**22.4**

set is {0,4,7,10}, and the chord dictionary is restricted to triads only, then the hand-annotated chord is reduced to the equivalent triad, {0,4,7}, and this pitch class set is used as the comparative chord label for the chord generated by the method.

The chord accuracy results for the four chord dictionaries are shown in Table 6.20. The accuracy data gives the percentage of matches between the output of the *Harman* algorithm and the hand-annotated data, which is reduced to match the chord dictionary of *Harman*. The results demonstrate that the largest chord dictionary has the lowest accuracy;

progressively smaller dictionaries showed better results. The trend is visually apparent in Figure 6.23, which graphs the percentage match results for the corpus. As can be seen from the graph, the method performs reasonably well in conjunction with the triadic chord dictionary, with accuracy levels progressively falling off with each increase in the number and complexity of chord templates. The average accuracies across the corpus for the four dictionaries are 63.0% (triad dictionary), 47.2% (triads and sevenths dictionary), 30.7% (triads 7ths and 9ths dictionary), and 22.4% (all templates). The percentage of chord labels in the ground truth that had to be reduced to match the maximum chord length contained in the dictionaries are 35.2% (triads), 3.5% (triads and 7ths dictionary), and 2.1% (triads 7ths and 9ths dictionary). A significant weaknesses of the *Harman* method is the inability to distinguish passing notes and exclude them from consideration. The preludes which contain a greater quantity of stepwise movement and fewer arpeggiated figurations, drop off in accuracy quite markedly with the introduction of additional chords to the chord dictionary. This can be seen particularly in preludes with a large degree of stepwise motion in the voices. Compare, for example, the accuracy levels for Prelude 2 (Table 6.20), which features many neighbour note configurations (please refer to Figure 5.3). From the triad dictionary to the sevenths dictionary the result drops by 47.3%, from an accuracy of 77.0% to a low 29.7%. The problem relates to the number of pitches present in the chord dictionary templates. If the templates are restricted to triads, ornamental note configurations have less impact as there are no templates in the dictionary containing all or most of the pitches in the presented group. Issues start to arise when the template sizes are increased, for example to four or more pitches, because the algorithm is able to account for a greater number of the presented pitches within a single template. The preference rules are often not able to satisfactorily reduce multiple possible options to the correct underlying harmony. The situation is subtle: in some cases the presented pitches represent an ornamented triad; in others, they represent an extended chord. If both templates produce the same score, it is difficult to identify the correct choice without additional contextual information. The result for prelude 2 therefore falls further

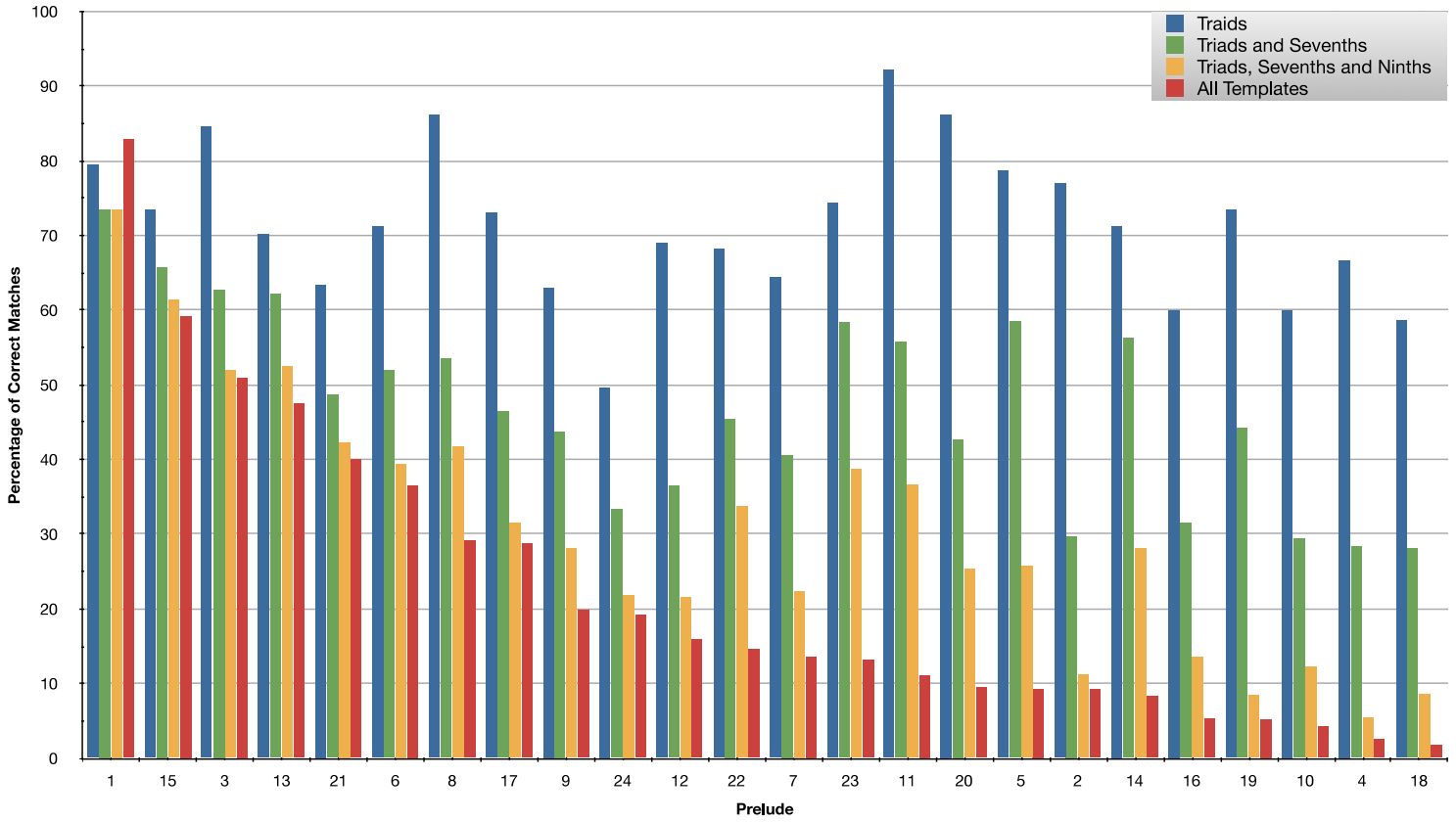


Figure 6.23: Accuracy levels of *Harman* method, Pardo and Birmingham [2002], in relation to four different chord dictionaries. Accuracy deteriorates as the quantity and complexity of chord templates defined in the dictionary increases.

with each successive addition to the defined chord range, reaching a very low result of 9.2% in conjunction with the largest chord dictionary. By this point, the accuracy level averaged across all of the preludes has dropped to 22.4%. The *Harman* algorithm requires the addition of some means of differentiating between inessential tones and structural chord tones to be able to successfully process complex notegroups in conjunction with more comprehensive chord dictionaries. The distribution of melodic steps per prelude in relation to the chord accuracy results of the *Harman* method using the triads and 7ths chord dictionary can be viewed in Figure 6.24. The correlation coefficient is 0.4. As a final experiment, the *Harman* method is run again in conjunction with the maximum dictionary size containing all templates. This time, all notes classified as inessential by the methods described earlier in this thesis, (neighbour notes, passing notes or pedal notes), are omitted from consideration by the algorithm. The removal of inessential notes from the groups demonstrates a marked improvement in results - the average accuracy across the corpus is lifted from 22.4% to 40.2%.

6.5.2 Matching Best Note Combination Sequences to Chord Templates

The output *BNC* sequences obtained using the combined score (see section 6.4.6) are matched to all four chord dictionary types to obtain a chord classification per combination using the evidence method of Pardo and Birmingham [2002] as described in the previous section.

In order to find a method of more effectively capturing complex chord designations even when chord factors are missing or obscured by additional pitches we evaluate three different methods of weighting the chord dictionary data. The weights are used to emphasize the scale degrees which are most indicative of a particular chord configuration and de-emphasize the most common note omissions from complex chords. A basic vector contains only 0 and 1, where 1 represents a chord note, and 0 represents a non-chord note. A C major chord template, {0,4,7} has a basic vector of [1, 0, 0, 0, 1, 0, 0, 1, 0, 0, 0, 0]. A second arrangement (*profile*

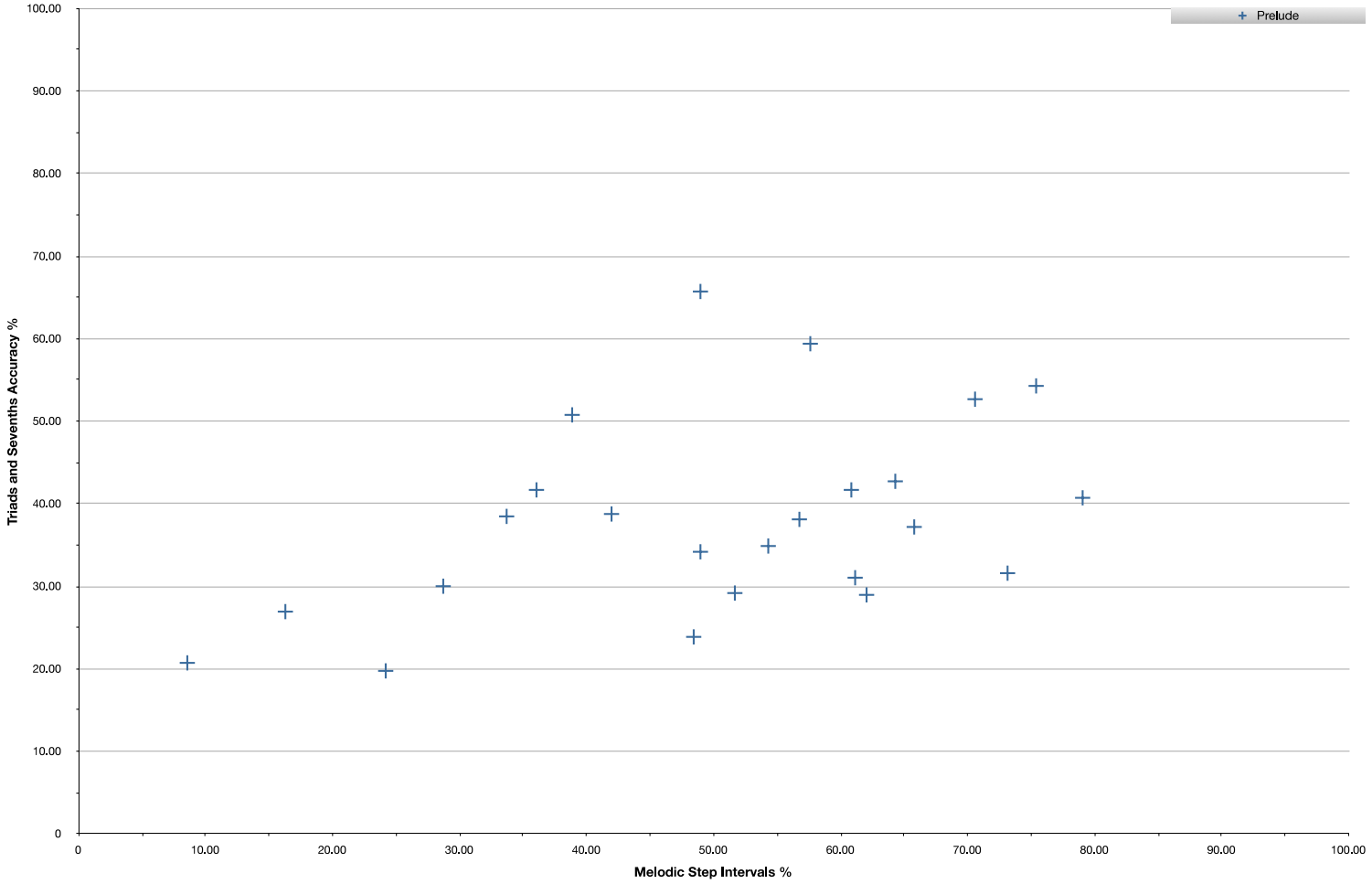


Figure 6.24: Scattergraph of distribution of melodic intervals against the chord accuracy values for the triads and 7ths chord dictionary using the *Harman* method. The correlation coefficient is 0.4.

weights) distinguishes between important and less important chord notes by increasing the weight of important chord factors to a rating of 2. In all cases it is assumed that the outer interval defines the chord most strongly, for example in the case of a seventh chord, it is the outer interval of the seventh that is most indicative. Selected inner notes are also rated as important, consequently, a triad weight profile is the root and 5th, a 7th profile weights the root, 3rd, 7th, a 9th weights the root, 3rd, and 9th, and 11th weights the root, 7th, 11th and a 13th, which typically omits the 5th, 9th, and 11th, weights root, 3rd, 7th, and 13th. The profile weighted vector for a C major triad is [2, 0, 0, 0, 1, 0, 0, 2, 0, 0, 0, 0] and for a G7 ({7, 11, 2, 5}) [0, 0, 1, 0, 0, 2, 0, 2, 0, 0, 0, 2]. In a final type of weighting (*root weights*), the profile weighted chord vectors are modified to stress the root note with a value of 3. When compared to a weighted template vector, if a presented pitch class matches a weighted template element, the match count is incremented by the template weight value. Therefore a C major triad would be represented by a root weighted vector of [3, 0, 0, 0, 1, 0, 0, 2, 0, 0, 0, 0], and the G7 chord by [0, 0, 1, 0, 0, 2, 0, 3, 0, 0, 0, 2]. The match count of a pitch class set S to a template $W = [W_0, ...W_{11}]$ is given by $\sum_{i \in S} W_i$, while the missed count is given by $-\sum_{i \notin S} W_i$, and the template match score is $\sum_{i \in S} -\sum_{i \notin S} W_i$. For example, a pitch class set of {0, 3, 7}, matched to the C Major weighted template, would produce a match count of 4 (0 and 7), and a missed count of -1, (3), resulting in a total template match score of 3.

Prior to presenting the final template match results, we analyse the nature of chord label designation of complex beat segments in more detail. For any one beat segment we compute how many of the chord tones (CT) in the hand-annotated chord are actually present, (i.e. are performed), in the input segment (henceforth referred to as the *input set*), how many of the chord notes are missing (CT-) from the *input set* prior to processing, and similarly, how many non-chord notes (NCT) are present in the *input set* when compared to the content of the hand-annotated chord. Subsequently, following processing by the combinations algorithm, we again count how many of the non-chord notes (NCT) remain in the BNC and

how many chord notes (CT). Clearly, the fewer chord tones (CT) that are actually present in the *input set* and by definition therefore, the BNC, and the greater the quantity of NCT present in either group, the more challenging it is to accurately obtain the desired chord designation from the processed group (BNC). The difference between *input set* and BNC in terms CT, CT-, and NCT can also tell us how effective the combinations method has been at a) correctly capturing structurally important chord tones from what was available at the input, and b) removing non-chord tones from the input segment to produce a pitch combination that is closer to the desired chord tone group.

Consider, as an example, the chord template match results for the segment in bar 8, beat 1, in Prelude 7. The hand-annotated chord is {10, 2, 5} (B♭ major triad), and the unique set of pitch classes contained in the *input set* are {10, 2, 0, 3, 9}. Following processing by the combinations algorithm, the BNC contains the pitch classes {10, 2} i.e. pitch classes {0, 3, 9} have been removed. All of these are represented in the hand-annotated data as NCT.

For this example, when compared to the hand-annotated chord and to each other, are as follows:

1. % of annotated chord tones (CT) present in the *input set* 66.6
2. % of *input set* non-chord tones (NCT) 60.0
3. % of BNC chord tones (CT) 66.6
4. % of BNC non-chord tones (NCT) 0
5. % non-chord tones (NCT) removed as a result of processing (i.e. the difference between the *input set* and the BNC) 100.0

The three types of measures (tertian, note importance and combined), prior to the application of the reduction rules, produce the following sets of equal scoring chord template matches:

1. Tertian Score - Top Scoring Template: {9, 0, 3}
2. Note Importance Score - Top Scoring Templates: {10, 2, 5}, {7, 10, 2}, {2, 6, 10}, {6, 10, 2}, {10, 2, 6}

3. Combined Score - Top Scoring Templates: {10, 2, 5}, {7, 10, 2}, {2, 6, 10}, {6, 10, 2}, {10, 2, 6}

As both the note importance score and the combined score contain more than one template match, a single template is chosen via a preference rule reduction process (detailed below). The final resulting templates are as follows:

1. Tertian Template: {9, 0, 3}
2. Note Importance Template: {10, 2, 5}
3. Combined Score Template: {10, 2, 5}

Table 6.21 gives statistics detailing the proportion of segments containing *NCT*, *CT*- and where the quantity of *NCT* is greater than *CT*, for *input set* data and for the processed *BNC* segments across the corpus. Averages for the corpus are shown at the bottom of the table.

The reduction of segments containing *NCT* in the *input set* data from an average of 75.8%, down to an average of 30.9% for the *BNC* data, shows that the algorithm is successfully removing *NCT* from more than half of the input segments. A smaller quantity of segments (14.3%) have legitimate chord tones removed during processing; the average quantity of segments missing chord tones (*CT*-) moves up from 28.9% in the *input set* data to 42.1% for the *BNC* data. The increase is relatively small when compared to the quantity of segments where *NCT* are successfully removed. Almost 20% of input data segments contain an equivalent number or higher of *NCT* compared to *CT*. This value is reduced to just over 10% of segments in the processed *BNC* data.

The data evidences that the production of multiple equal scoring chord templates from the *BNC* chord matching method is a significant problem with 41.3% of segments across the corpus resulting in multiple possible chord symbol choices. This is an area for improvement in future work; having to select a chord label from a number of possible options increases the potential for erroneous labelling.

To assess the effectiveness of weighting the chord templates on the

Table 6.21: Statistical comparison of ground truth chord tones in relation to input segment tones and processed note group tones (*BNC*) across the corpus. The columns entitled *NCT* give the percentage of segments containing non-chord tones. The columns headed *CT-* give the percentage of segments with missing chord tones. The columns heading *NCT>CT* give the percentage of segments where the number of non-chord tones is equal to or greater than the number of chord tones. The *Multiple templates* column refers to the production of more than one possible chord template match. Columns 2-4 give data about notes in the input segment, columns 5-8 show statistics for the *BNC*. All values are expressed as a percentage of the total number of segments in the sequence containing these features.

Prelude	**Input Segments**			**BNC Segments**			**Multiple**
	NCT	***CT-***	***NCT>CT***	***NCT***	***CT-***	***NCT>CT***	**Templates**
1.0	10.7	14.3	0.0	5.7	15.7	0.0	20.7
2.0	91.5	34.2	19.1	19.7	40.1	9.9	43.4
3.0	45.2	21.2	21.2	20.2	32.7	5.8	35.6
4.0	98.7	10.3	46.2	57.7	41.0	25.6	42.3
5.0	93.6	52.9	10.0	36.4	68.6	12.9	60.0
6.0	59.6	9.6	14.4	37.5	28.9	11.5	35.6
7.0	85.7	29.3	21.1	42.1	52.1	14.6	49.6
8.0	63.3	20.0	9.2	24.2	26.7	7.5	40.0
9.0	78.1	36.5	21.9	37.5	46.9	12.5	51.0
10.0	98.2	32.9	28.7	44.5	51.2	14.6	59.8
11.0	98.6	6.9	4.2	16.7	16.7	1.4	13.9
12.0	81.8	18.2	19.3	45.5	39.8	23.9	40.9
13.0	54.2	37.5	10.8	21.7	38.3	9.2	40.8
14.0	96.9	35.4	11.5	14.6	41.7	3.1	56.3
15.0	41.5	29.0	9.2	21.1	31.6	3.3	16.5
16.0	97.4	14.5	23.7	56.6	42.1	21.1	39.5
17.0	69.7	43.9	13.6	24.2	50.8	7.6	42.4
18.0	58.6	49.4	7.5	22.4	51.7	5.2	44.3
19.0	95.8	57.3	29.2	32.3	71.9	17.7	55.2
20.0	90.5	11.9	22.6	28.6	32.1	7.1	29.8
21.0	57.5	16.3	15.0	36.3	45.0	10.0	42.5
22.0	86.5	24.0	12.5	34.4	34.4	3.1	29.2
23.0	86.8	38.2	7.9	23.7	50.0	2.6	39.5
24.0	79.3	49.5	16.0	38.8	60.6	17.0	61.7
Average	**75.8**	**28.9**	**16.4**	**31.0**	**42.1**	**10.3**	**41.3**

chord match process, we evaluate the impact of the different chord template representations. For each dictionary weight format, the following results are counted: the number of times multiple equal scoring templates are generated per single input segment per prelude, the number of times

Table 6.22: Impact of weighted chord templates on accuracy and multiple template generation.

Weight Type	% Multiple Top Scoring Templates	% Ground Truth Chord in Multiple Match	% Matched Single Results
Basic (no weight)	25.2	62.7	60.7
Profile Weights	28.2	61.2	57.0
Tonic Weights	29.1	57.5	55.5

a list of multiple top scoring templates contains the desired ground truth template, and, where there is a single chord template result for a segment, the number of times this single chord template matches the ground truth chord. As can be see from the results shown in Table 6.22, and contrary to expectation, the non-weighted chord templates produce the best results for all three categories measured.

The final stage is to reduce all multiple top scoring template lists per segment to a single chord choice so that a sequence of individual chord labels is produced for each prelude. For each set of multiple templates, preferred templates are selected by computing the closest match between the template pitches and the note combination pitches. This is calculated using mathematical set theory. In mathematics, the *intersection* of two sets A and B is the set that contains all elements of A that also belong to B, or vice versa. The *symmetric difference* takes set A and B and computes a new set containing elements in either A or B but not in both. The combination of these two membership tests can be used to ascertain the degree of pitch commonality between a template set and a note combination set.

For example, given the *BNC* note combination [1, 8], (call it A) and two templates [1, 4, 8] (*B1*) and [1, 4, 8, 11] (*B2*), the *intersection* of A and *B1* is the same as A with *B2*, (set [1, 8]), however the *symmetric difference* between A and *B1* is a smaller set ([4]) than the *symmetric difference* between A and *B2* (set[4, 11]), and for this reason template *B1* is selected as the one most closely matching set A. The membership test of pitches method proves to be an effective way of narrowing down a list

of choices from the templates. Should more than one top scoring template remain, the options are reduced again in accordance with the chord profile probability table of Pardo and Birmingham [2002], in which chords are preferred in accordance with their interval profile as follows: major, dominant 7th, minor, diminished 7th, half-diminished 7th, diminished triad. For example, should the list contain a dominant 7th and a minor chord, after the application of the preference rule, only the dominant 7th chord would remain. Finally, due to some multiples containing different inversions of a diminished 7th chord, a single chord choice is preferred based by matching the root note of the chord with the lowest pitch of the note combination. This combination of rules reduces the options to a single choice for each note group for the corpus.

6.6 Results and Discussion

We have shown that identifying the underlying harmony in ornamental keyboard music is a challenging task both for the human annotator and for automatic methods of chord recognition. Variations in harmonic rhythm, melodic processes, and note emphasis techniques; sparse textures from which critically defining chord notes are missing; ornamentation featuring contrasting chord and non-chord tones; and above all, ambiguity of both key or chord in some segments, renders this an extremely difficult task. It would be useful to have an upper limit of accuracy based on human annotations of this or another corpus of common practice musical works, but to obtain this kind of measure a single musical corpus would need to be annotated by a number of human annotators using the same beat segmentation and adopting the same approach and methodology. The level of deviation between the resulting annotated sets could be measured. However, there is no data available for this.

The overall results for the combinations (*BNCs*) matched to chord templates using all four chord dictionary types are shown in Table 6.23. The average result for the corpus for the triads dictionary is slightly below that of the *Harman* result at 59.7%, however our method reduces all chord options to a single chord choice while *Harman* is left with multiple options,

Table 6.23: Chord accuracy results using the combinations method with four different chord dictionaries. Chord options are reduced to a single chord option in the method.

Prelude	**Triads**	**Triads and 7ths**	**Triads, 7ths and 9ths**	**All Templates**
1.0	79.3	89.2	86.3	86.3
2.0	62.3	62.0	62.0	62.0
3.0	69.2	65.4	64.4	64.4
4.0	47.4	40.8	39.0	36.8
5.0	49.6	43.1	41.6	41.6
6.0	61.5	56.7	56.7	56.7
7.0	55.2	45.0	44.6	44.9
8.0	61.7	67.5	66.7	66.7
9.0	61.5	53.1	53.1	52.1
10.0	49.1	41.0	40.4	40.0
11.0	77.8	81.9	81.9	81.9
12.0	52.9	51.7	49.4	49.4
13.0	63.3	65.0	64.2	64.2
14.0	63.2	62.1	62.1	61.1
15.0	73.7	69.1	69.1	69.1
16.0	47.4	35.5	35.5	35.5
17.0	62.1	57.6	56.8	56.8
18.0	62.1	58.6	57.5	57.5
19.0	41.7	38.5	38.5	37.5
20.0	63.9	63.9	62.7	62.7
21.0	61.3	48.8	45.0	45.0
22.0	58.5	60.6	57.5	58.5
23.0	61.8	54.0	51.3	51.3
24.0	46.8	42.5	40.9	37.6
Average	**59.7**	**56.4**	**55.2**	**55.0**

any one of which could be the correct template, thus the evaluation of our method is stricter. The performance of the method in conjunction with all of the four dictionaries is highly consistent across all dictionary types: averages for the four dictionaries are 59.7%, 56.4%, 55.2% and 55.0%, thus there is only a 4.7% deterioration of results from the simplest Triads dictionary to the most comprehensive All Templates dictionary. The trend is a significant improvement over that evidenced by the *Harman* method. Considerably higher levels of accuracy are achieved for individual preludes containing a high proportion of passing notes and ornamental non-chord tones in comparison to the prototype method.

Harman has been found to be a useful labelling method if simple triadic structures are all that is required; however, for the purposes of in-depth automatic music analysis and more profound style characterisation, the fundamental triad types are considered to be insufficiently descriptive. In addition, the method features no processing of the digital data to identify inessential notes prior to labelling, consequently the effectiveness of the method is significantly impacted by the presence of such notes in input groups, and yields much lower accuracy results in conjunction with ambiguous data. The method also progressively drops off in accuracy in relation to the addition of complex chord types to the chord dictionary, consequently without modifications the method is not suitable for the generation of information rich harmony labelling, particularly in conjunction with complex corpuses. Removing inessential notes from the input group was found to improve the effectiveness of the *Harman* method by almost 20% when used in conjunction with the All Templates chord dictionary. The method generates large numbers of multiple top scoring chord definition templates, thus extensions to existing preferences rules for the purposes of final chord selection are necessary.

This chapter also details significant work in the automatic processing of digital scores in order to access the kind of note features that would be used by a human analyst when annotating music with chord labels. Prior to obtaining note features, the score data is segmented into temporal beat segments corresponding to compound or simple beats, and a novel musical voice separation method, a further pre-requisite processing stage

necessary for the identification of linear note features (such as passing notes and neighbour notes), demonstrates a high level of accuracy when compared to MIDI ground truth data. A new voice thresholding approach estimates the number of musical voices in the score and successfully counters the impact of dense chords on overall voicing values. The chapter goes on to detail a novel method of passing note identification based on linear voice stepwise movement and the tertian intervallic relationships of classified notes to surrounding notes. The interval relationships are used to ascertain the degree of certainty of the passing note classification. Notes of duplicate pitch within the same segment are subsequently re-classified to align all matching pitches. The passing note method is not tested directly on ground truth data, however the subsequent calculation of the percentage of non-chord tone removal from input data shows that 60% of the input sets with non-chord tones have the non-chord tones successfully removed by the combinations method (Table 6.21).

An important point of note is that this work addresses two specific categories of inessential notes *only*: neighbour notes, and passing notes situated in a stepwise series of three. There are many other categories of inessential notes; these are not covered by the work described here. Consequently, it is not possible to ascertain the extent to which the 40% remainder of input sets still containing non-chord tones is attributable to errors in the detection of non-chord tones, or the extent to which it is attributable to the presence of categories of inessential notes that are not accounted for.

The combinations method makes no prior decision regarding the number of notes present in any one segment that form part of a chord or best combination; all distinct note combinations in all segments are submitted to the algorithm as potential chord combinations. Note features, including metrical and durational emphasis, and inessential note classification, are used to derive a measure of individual note importance within the context of the segment, and the tertian arrangement potential of each subset of notes is also calculated. The two measures are combined to produce a final combined measure, thus generating three distinct sequences of top scoring best note combinations per prelude. The three sets of sequences

are compared to hand-annotated ground truth data. The tertian arrangement of notes produces a more accurate series of note combinations than the groups produced by note importance features, however the generation, interaction, and scoring of these features is complex and open to further experimentation. Combining measures of note importance with the scoring of tertian arrangements produces the best overall results for the dataset. Segmenting some of the preludes at the compound beat level is shown to produce situations of tonal ambiguity that are not resolvable by either the tertian arrangement or note importance rules used in this work; the method attempts to account for as many pitches as possible within a single label, and group subdivision is not attempted. Re-segmenting the data of a compound beat prelude to a smaller beat level and matching this to a unique set of annotations evinced considerable improvement in results, evidencing that the level of harmonic movement was often at the smaller beat level, and that accessing the harmonies at the reduced durational level was possible by the chord recognition method described in this chapter.

An observation from the work presented in this thesis is that the combination of inessential notes (or *NCT*) with omitted chord tones *CT-* presents significant difficulties. Figure 6.25 plots the accuracy data of the combinations method in conjunction with non-weighted chord templates compared to the proportion of *NCT* and *CT-* (missing chord tones) in the *BNC* segments. The graph visually demonstrates a relationship between the presence of spurious or conflicting tonal elements in the data, missing or omitted chord notes, and the capability of the software to correctly identify the underlying harmony. Accuracy levels significantly drop off as the quantity of positive chord evidence reduces and the proportion of alternative tonal elements increases. In some cases features of articulation, such as metrical emphasis, duration, registral emphasis and pitch repetition may be sufficient to overcome the conflicting chord information presented by missing chord notes or the presence of ornamental non-chord notes.

In some cases however, the features of articulation captured by the methods described do not cause a removal of non-chord tones, and this

presents a significant problem to the automatic chord identification methods explored.

Despite considerable success in the removal of non-chord tones from the best combinations (*BNCs*) segments, matching the note content of the *BNCs* to chord dictionaries in order to obtain a single final and accurate chord label per segment remains challenging. The greater the number of non-chord tones, and the fewer defining chord tones (i.e. chord tones that have been omitted), the harder it is to access the underlying harmony. The problem is exacerbated by the inclusion of complex chord types such as 9ths, 11ths and 13ths in the template dictionary, often resulting in erroneous chord matches if the presented notes contain conflicting tonal content. For the advancement of systematic musicology research however, resolving the automatic capture of extended chord types is mooted to be a necessity. An alternative method of matching notes to dictionary templates is suggested for future research, (see next chapter), representing complex chords by their common musical pitch articulation, rather than by including every one of the component tones of the chord.

An improvement in the level of sophistication to automatic harmony processing has a great deal of potential for systematic musicology. Enhancements in this area may lead to the production of high quality harmony data that could be used in the automatic analysis of large corpuses. In addition, although this work has been tested on a corpus in which the use of extended chords is relatively limited, in the interests of general applicability, richness of information, and facilitation of systematic musicology research, automatic methods must be capable of accessing vertical sonorities beyond those of basic triads and sevenths. Once this has been achieved it will become possible to access characteristic high level aspects of a composer's harmonic language potentially producing novel insights that are supported by quantitative data.

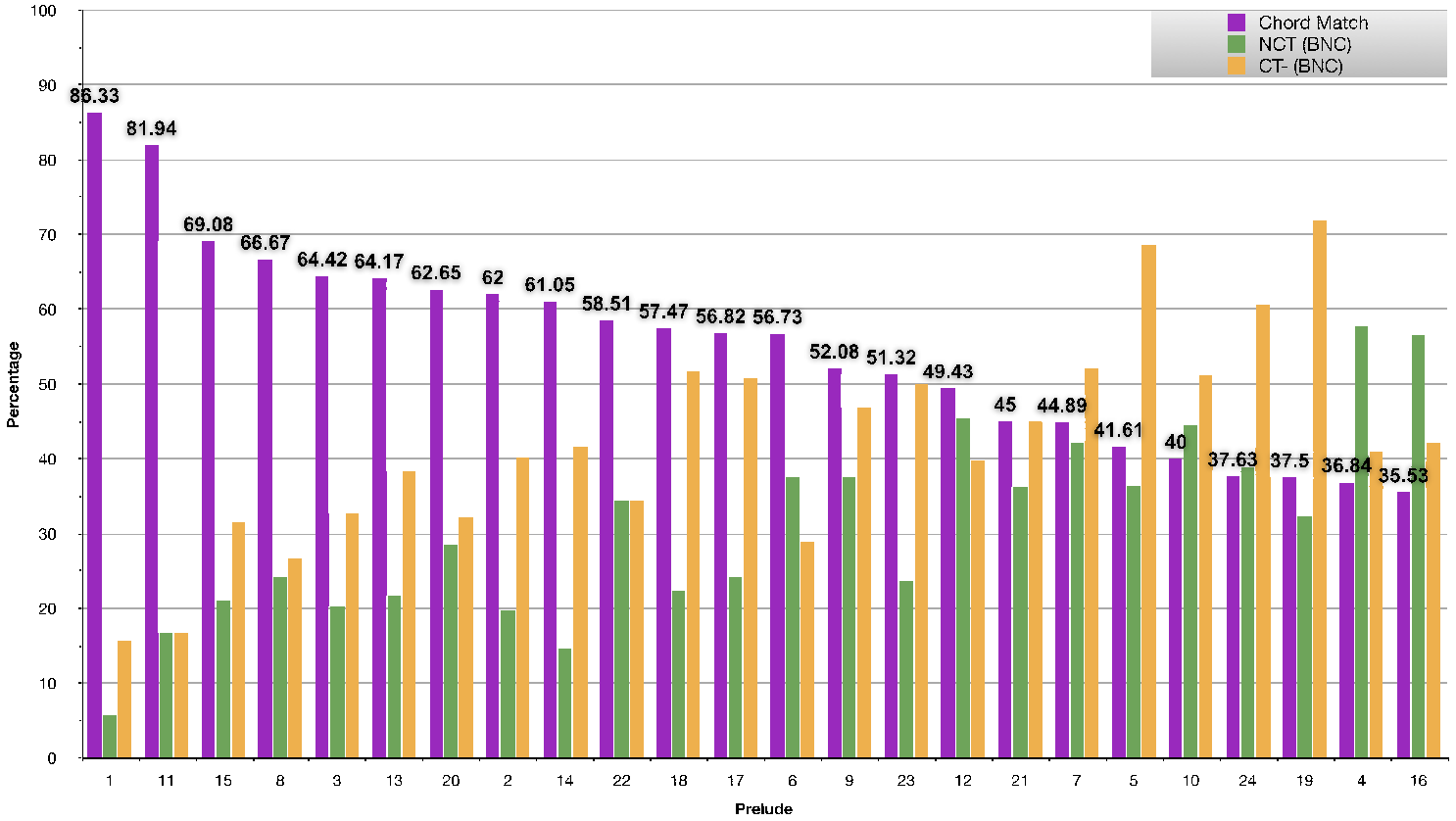


Figure 6.25: Chord accuracy results of note combinations using the All Templates dictionary compared to the percentage of *NCT* and *CT-* for the *BNC* segments per prelude. The graph demonstrates the impact of non-chord tone or missing chord tone elements in the segments on chord match accuracy, with accuracy levels decreasing as the proportion of *NCT* and *CT-* increases.

Chapter 7

Conclusions and Future Work

This thesis commenced by stating that the aim of this research was to bring the disciplines of musicology and computer science closer together. Three main aims were identified, the primary focus of which was an improvement in computational methods of extracting harmony information from digital corpuses. Systematic musicology offers immense potential for novel exploration of music producing quantitative results, but for systematic approaches to deliver the kind of high quality and transparent information about complex music such as that of the Baroque period, a great deal of work remains. This chapter summarises the discoveries made during the course of this doctoral research, and suggests developmental areas for future work in the field.

7.1 Conclusions

Chapter 1 highlights the language and approach of the discipline of musicology, and defines the differences between critical analysis (the close study of a single work), and style analysis (the study of commonalities across a body of works), in music. The writings of Deliege and Cambouropoulos are discussed with reference to definitions of a 'musical surface'; in particular, the idea that human listeners make sense of music by abstracting complex musical structures, or 'wholes', from the sequences of note events they hear, and that this concept needs to be moved across into computational work in order to make any real progress in computational musicology. A motivation of the work presented in this thesis is to pursue the idea of abstracting more meaningful musical constructs, thus expediting automatic

harmonic analysis.

Chapter 2 presents an overview of core concepts in musicology, providing details of the theoretical foundations on which the research is built. Topics covered include counterpoint, harmony and metre, and the relationship of these to broader topics such as musical style. Music theoretic and analytical theories that have influenced the approaches adopted in this research are outlined.

In Chapter 3 research relating to the computational abstraction of musical constructs and harmony are reported. The most successful approaches to musical voice separation and automatic chord symbol extraction, in terms of accuracy levels or popular adoption in the community, are given particular attention due to their influence on the work presented here. The chapter highlights the difficulty in the abstraction of core musical concepts such as chords, key and voicing from digital data, and attests that producing high quality software to analyse music continues to be a challenging problem.

Chapter 4 describes key and modulation detection from automatically generated chord sequences using hidden Markov models (HMMs). One of the most interesting results of the work was the accuracy of the results obtained from audio data compared to the MIDI data; errors which occurred during transcription were progressively smoothed out during chord and key modelling, ultimately resulting in comparable levels of key output accuracy and suggesting the potential for this type of digital data for future research as transcription methods continue to improve. Exploiting the framework of the models to test the effectiveness of Krumhansl's perceptual chord and key data against values elicited heuristically from the harmonic theory of Arnold Schönberg, we found that values linked to music theoretic principles consistently produce the most accurate key change results, both in terms of determining the precise moment of key change - a widely acknowledged problem in this type of work - and in terms of pinpointing the correct key, when compared to hand annotated ground truth data. The work expands previous work in this area by including larger sets of chord symbols; models testing perceptual data necessarily symbolise the 48 major, minor, diminished and augmented triads, however further

models add the full range of seventh chords, as defined by Schönberg, to this basic set. The models based on the more complex observation data produce more equivocal results, partly due to the imbalance between the number of chords defined as being part of the minor key compared to the number of chords representing the major key (i.e. considerably more chords are defined as symbolising the minor key), weighting the models towards major key outputs. In addition, the more equivocal results of seventh model suggests that extended chords increase tonal ambiguity in contrast to simply triads which are able to define key more clearly. Significant deviation in key outputs between the various models highlights areas of complex or ambiguous harmony in the data set. Overall, all of the models produce key data of sufficient accuracy to obtain measures of modulatory type, frequency and key distance.

The final two chapters of this thesis present work to abstract the underlying chord structure of the complex keyboard music of J. S. Bach, by removing inessential notes and identifying tertian pitch relationships. To our knowledge this is the first attempt to classify inessential notes in ornamental music and use this information to extract a broad range of chords from an intricate test corpus.

An important contribution of this work is the annotation of ground truth chord data for J. S. Bach's first twenty four preludes of the Well Tempered Clavier, Book One. Chapter 5 describes the process of hand annotating the dataset, subsequently used to evaluate the output sequences of the structural note methods. Prior to this, no complete sets of chord or key annotations exist for the test corpus; the partial annotations by Riemann provided a valuable source of reference [Riemann, 1890]. A further contribution is the extension of Chris Harte's chord annotation syntax [Harte, 2010], to facilitate the annotation of western classical music. Modifications include: allowing more than one permissible chord label; the representation of pedal notes; and additional shorthand labels for extended chords. A software parser converts the chord labels to pitch class set format, optionally containing pedal note information.

The data processing required to reach the stage of being able to test

a structural note algorithm based on abstracted note features is considerable. A method to segment into linear voice groups extends the approach of Chew and Wu [2005] and yields a high degree of accuracy. Novel additions include an alternative method of data segmentation, repeat parsing to clarify ambiguous linear note connections, the inclusion of dense chords (removed by other algorithms due to their negative impact on results), and a thresholding method to ascertain optimum voicing. A novel aspect of the work is the computation of a measure of contextual note importance from note features that aims to mirror the musical articulation. Following rigorous feature definition, contour, metrical strength and inessential note classifications are determined from the digital data. Although the passing note method is not tested empirically, subsequent evaluation of the chord results demonstrates that a high proportion of inessential notes are successfully removed from input groups (see below).

The method intentionally avoids predetermining the note membership of structural note groups, and using a brute force approach scores all of the distinct note combinations of an input group. The first scoring method is based on note features (such as metre or note classification), the second method computes the tertian chord potential of each subset. The two types of scores are then combined to obtain a final third combination score (and associated note combination). In each case, multiple top scoring combinations of notes are reduced to a single choice via musically informed preference rules. The result is three unique sequences of best note combinations (*BNCs*) per prelude. The combination of both note features and tertian arrangement potential, (third type of measure), produces the best series of *BNCs* when compared to hand annotated data, with an average level of accuracy of 60.8% across the corpus.

Comparison of the notes in the hand annotated chord data, the notes presented in the input data, and the notes contained in the *BNCs*, affirms that the methods are successful in the removal of non-chord tones from segments *NCT*, with 60% of segments having *NCTs* removed. A comparatively smaller quantity of chord tones *CT* are also erroneously removed during processing. Matching the *BNC* sequences to four different types of chord dictionary produces consistent accuracy levels across the four

dictionary types, (59.7%, 56.4%, 55.2% and 54.9% for each of the dictionaries from simplest to most comprehensive), clearly outperforming the prototype *Harman* method in this respect. The work has produced quantitative data evidencing that the relative proportions of *NCT*, *CT* and missing or omitted chord tones *CT-* in input segments and *BNC* directly influence chord abstraction accuracies. The prototype method clearly exposes the problem of allowing inessential notes to be present in input sets to any chord algorithm for this type of data: the method generates large quantities of multiple equal scoring templates per segment, and reveals progressive and significant deterioration of accuracy levels directly in line with the increase in size of chord dictionary, (specifically, averages of 63.0%, 47.2%, 30.7% and 22.4% from the smallest to the largest dictionary). Individual preludes give particularly poor results when they are segmented at the larger beat level, (for example prelude 18), revealing that the method is weak at abstracting chord designations when more than one legitimate chord is present within a segment. The removal of notes from input sets designated as inessential by the methods described in this thesis are shown to lift the final *Harman* result from 22.4% to 40.2%.

Our work proves that correctly differentiating between inessential and chord notes in complex music is an imperative component of any approach to musical harmony in complex corpuses; more so when the aim is to access diverse, accurate and information-rich harmonic descriptors. The work presented in this thesis shows that the development of computational methods that move beyond the superficial stream of notes at the surface level of a score are vital to the advancement of the state of the art in computational musicology and MIR. The work is possibly relevant to automatic genre recognition and music recommendation systems. Some ideas for future work in this respect are outlined in the following section.

7.2 Future Work

Several areas for future work have presented themselves during the course of this doctoral research, these are now detailed as follows.

7.2.1 Musical Voicing

The intractability of the problem of musical voicing lies less with the difficulty of producing algorithms capable of accurately voicing music, than with difficulty of arriving at a precise definition of what a musical voice actually *is*. The term has evolved from early 'a capella' vocal practice in which there are fixed number of (human) voices sounding a single note per voice, to later keyboard idioms, in which a single musical 'voice' may have many notes at any one time and in which the texture is constantly varying. The clear indication is that it is not valid to restrict a computational model to one note per voice, either from a compositional or perceptual perspective. In addition to this, there is the concept of voice-leading, another oft-used term with multiple interpretations. In Schenkerian analysis, voice-leading refers to higher level motivational and structural forces in the music, rather than note to note progressions at the surface level of part writing. A theoretical model representing an alternative viewpoint of musical voicing with the potential to separate out the individual voice elements of compound melody and access higher level voice-leading structures would therefore be a significant contribution to systematic musicology. Vertically clustering pitch groups would enable pitch range similarities to identify linear connections and thus indicate voice membership rather than the more usual method of pitch succession. An example of the proposed model can be seen in Figure 7.1. In the figure, the first one or two bars of the first four preludes is shown followed by the vertical stacking of notes per beat, disregarding the metrical position or duration of individual notes within the beat. The registral organisation of the notes within the beat is rendered immediately apparent. In the figure, note stems are used to indicate membership of one of a maximum of four voices. Although the examples shown in the figure do not signify the voicing of the entire work, the excerpts provide a useful basis for discussion in order to understand the possibilities of a vertical stacking method to discover how notes should be integrated vertically into a single musical voice.

7.2.2 Passing Note Identification and Chord Algorithm Improvement

Indirect testing of the passing note method indicates that it is successfully classifying a proportion of passing notes in the test corpus. To understand precisely how well the algorithm works in its current format, and to quantitatively test, research and improve the algorithm, the most important next step is to create the necessary ground truth data against which to compare the method. Such data will be drawn from the existing test corpus, but also from contrasting corpuses in order to produce a passing note algorithm that is as generally applicable as possible. A further important improvement on current work, is to incorporate note feature abstraction fully into the combinations algorithm, and to simultaneously assess both note features and tertian interval relationships. This has the potential to raise the accuracy levels described in this thesis. Individual note features, for example, metre, may be systematically tested, to ascertain the degree of influence each feature has on chord results. Consideration needs to be given also to the issue of missing chord tones, which could potentially be addressed during combinations processing, or subsequently via different types of chord templates.

7.2.3 Refinement of Hand Annotated Data for Use by the Community

The hand annotated chord and key data listed in chapter 4 and the chord data in chapter 5 was created by a single annotator. Existing sources of harmony data for the test corpuses were referred to as far as possible during the process, however, to be fully assured that the datasets are robust and accurate, in particular with respect to the more complex and equivocal Bach preludes, an important future goal is to involve other musicians in the production and/or verification of the data. At present, parallel key data exists only for a small quantity of the preludes test corpus; complete sets of robust chord and key annotations for both chorales and preludes by J. S. Bach will be an invaluable contribution to the MIR community,

enabling novel research into harmony algorithms. The current set of extended chord definitions in the syntax will also be extended to produce a comprehensive set of chord classifications representative of later musical periods and jazz.

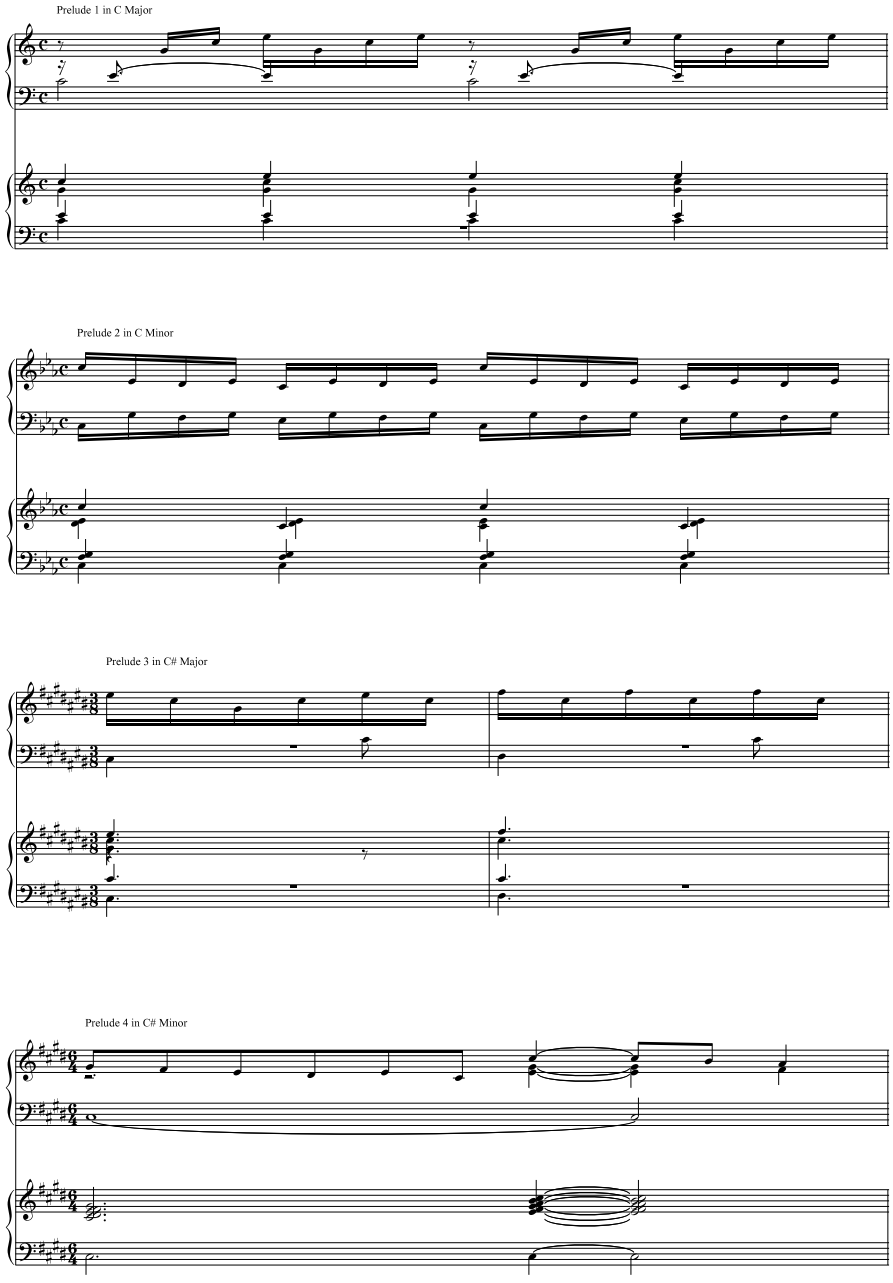


Figure 7.1: Vertical stacking of notes per beat to obtain voice groupings.

Bibliography

B. Alegant. Cross-partitions as harmony and voice leading in twelve-tone music. *Music Theory Spectrum*, 23(1):1–40, 2001.

A. Anglade and S. Dixon. Characterisation of harmony with inductive logic programming. In *Proceedings of the 9th International Conference on Music Information Retrieval (ISMIR 2008)*, pages 63–68, 2008.

A. Anglade, R. Ramirez, and S. Dixon. Genre classification using harmony rules induced from automatic chord transcriptions. In *Proceedings of the 10th International Conference on Music Information Retrieval (ISMIR 2009)*, pages 669–674, 2009.

A. Anglade, E. Benetos, M. Mauch, and S. Dixon. Improving music genre classification using automatically induced harmony rules. *Journal of New Music Research*, 39(4):349 – 361, 2010.

E. Antokoletz. Interval cycles in Stravinsky's early ballets. *Journal of the American Musicological Society*, 39(3):578–614, 1986.

W. Apel. *Harvard Dictionary of Music*. Heinemann Educational Books Ltd, London, 1970.

E. Benetos and S. Dixon. Polyphonic music transcription using note onset and offset detection. In *Proceedings of the 2011 International Conference on Acoustics, Speech, and Signal Processing*, pages 37–40, May 2011.

B. J. Blackburn. On compositional process in the fifteenth century. *Journal of the American Musicological Society*, 40(2):210–284, 1987.

A. S. Bregman. *Auditory Scene Analysis: The Perceptual Organisation of Sound*. MIT Press, Cambridge, MA, USA, 1990.

M. Bribitzer-Stull. The a-c-e complex: The origin and function of chromatic major third collections in nineteenth-century music. *Music Theory Spectrum*, 28(2):167–190, Fall 2006.

S. C. Brown. ic1/ic5 interaction in the music of Shostakovich. *Music Analysis*, 28(2/3):185–220, 2009.

E. Cambouropoulos. Voice and stream: Perceptual and computational modelling of voice separation. *Music Perception*, 26(1):75–94, 2008.

E. Cambouropoulos. How similar is similar. *Musicae Scientae*, 4B(7-24), 2009.

E. Cambouropoulos and C. Tsougras. Auditory streams in Ligeti's Continuum: A theoretical and perceptual study. In *The Fourth Conference on Interdiscplinary Musicology (CIM08)*, 2008.

P. Cathé. Harmonic vectors and stylistic analysis: a computer aided analysis of the first movement of Brahms' string quartet op. 51-1. *Journal of Mathematics and Music*, 2010.

E Chew and X Wu. Separating voices in polyphonic music: A contig mapping approach. In UffeKock Wiil, editor, *Computer Music Modeling and Retrieval*, volume 3310 of *Lecture Notes in Computer Science*, pages 1–20. Springer Berlin Heidelberg, 2005.

D. Conklin and M. Bergeron. Discovery of contrapuntal patterns. In *Proceedings of International Society for Music Information Retrieval (ISMIR)*, 2010.

N. Cook. Schenker's theory of music as ethics. *The Journal of Musicology*, 7(4):415–439, Autumn 1989.

N. Cook. Keynote talk: Towards the complete musicologist. In *Proceedings of the 5th International Conference on Music Information Retrieval (ISMIR 2005)*, 2005.

N. Cook. The psychophysics of harmony perception: Harmony is a three-tone phenomenon. *Empirical Musicology Review*, 1(2):106–126, Spring 2006.

J. D. Cuciurean. *A Theory of Pitch, Rhythm, and Intertextual Allusion for the Late Music of György Ligeti*. PhD thesis, State University of New York at Buffalo, 2000.

C. Dahlhaus. Harmony. `http://www.oxfordmusiconline.com`, 2007.

I. Deliege. Similarity relatons in listening to music: how do they come into play? *Musicae Scientae*, 4A:9–37, 2007.

D. Deutsch. Two-channel listening to musical scales. *Journal of the Acoustical Society of America*, 57:1156–1160, 1975.

D. Deutsch. Grouping mechanisms in music. *The Psychology of Music*, pages 99–134, 1982.

W. Drabkin. Part-writing. `http://www.oxfordmusiconline.com`, 2007.

D. Ferris. C. P. E. Bach and the art of strange modulation. *Music Theory Spectrum*, 22(1):60–88, 2000.

A. Forte. *The Structure of Atonal Music*. Yale University Press, 1973.

A. Forte and S. E. Gilbert. *Introduction to Schenkerian Analysis*. W. W. Norton and Company, 1982.

W. Frobenius. Polyphony. `http://www.oxfordmusiconline.com`, 2012.

T. Fujishima. Real time chord recognition of musical sound: a system using common lisp music. In *Proceedings of International Computer Music Conference (ICMC)*., 1999.

A. Gosman. Rameau and Zarlino: Polemics in the traité de l'harmonie. *Music Theory Spectrum*, 22(1):44–59, 2000.

M. Goto, H. Hashiguchi, T. Nishimura, and R. Oka. RWC music database: music genre database and musical instrument sound database. In *Proceedings of the 4th International Conference on Music Information Retrieval (ISMIR 2003)*, October 2003.

D. J. Grout. *A History of Western Music*. J. M. Dent & Sons Ltd, 1980.

M. Hamanaka, K. Hirata, and S. Tojo. ATTA: Implementing GTTM on a computer. In *Proceedings of the 8th International Conference on Music Information Retrieval (ISMIR)*, 2007.

C. Harte. *Towards Automatic Extraction of Harmony Information from Music Signals*. PhD thesis, Queen Mary University of London, 2010.

C. Harte, M. B. Sandler, S. A. Abdallah, and E. Gómez. Symbolic representation of musical chords: A proposed syntax for text annotations. In *Proceedings of the 6th International Conference on Music Information Retrieval (ISMIR 2005)*, pages 66–71, 2005.

R. Hillewaere, B. Manderick, and D. Conklin. Global feature versus event models for folk song classification. In *Proceedings of 10th International Society for Music Information Retrieval (ISMIR)*, 2009.

R. Hillewaere, B. Manderick, and D. Conklin. String quartet classification with monophonic models. In *Proceedings of 11th International Society for Music Information Retrieval (ISMIR)*, 2010.

P. Hindemith. *The Craft of Musical Composition. Book 1: Theory*. Schott, 1942.

D. Huron. Tone and voice: A derivation of the rules of voice-leading from perceptual principles. *Music Perception*, 19:1–64, 2001.

D. Huron. *Sweet Anticipation: Music and the Psychology of Expectation*. MIT Press, Cambridge, MA, USA., 2007.

A. Ishigaki, M. Matsubara, and H. Saito. Prioritized contig combining to segregate voices in polyphonic music. In *Proceedings of SMC Conference, Padova*. Sound and Music Computing, 2011.

J. Kilian and H. H. Hoos. Voice separation- a local optimisation approach. In *Proceedings of the 3rd International Conference on Music Information Retrieval (ISMIR)*, pages 39–46, 2002.

R. Kirkpatrick. *Interpreting Bach's Well Tempered Clavier*. Yale University Press, 1984.

P. B. Kirlin. Using harmonic and melodic analyses to automate the initial stages of Schenkerian analysis. In *Proceedings of the 10th International Conference on Music Information Retrieval (ISMIR 2009)*, 2009.

P. B. Kirlin and P. E. Utgoff. Voise: Learning to segregate voices in explicit and implicit polyphony. In *Proceedings of the 6th International Conference on Music Information Retrieval (ISMIR 2005)*, 2005.

P. B. Kirlin and P. E. Utgoff. A framework for automated Schenkerian analysis. In *Proceedings of the 9th International Conference on Music Information Retrieval (ISMIR)*, 2008.

C. H. Kitson. *The art of counterpoint and its application as a decorative principle*. Oxford at the Clarendon Press, 1907.

C. H. Kitson. *Elementary Harmony*. Oxford University Press, 1920.

S. Kostka and D. Payne. *Tonal Harmony*. New York: McGraw-Hill., 1984.

K. Kramer. The mirror of tonality: Transitional features of nineteenth-century harmony. *19th-Century Music*, 4(3):191–208, 1981.

L. Kramer. Chopin's rogue pitches: Artifice, personification, and the cult of the dandy in three later mazurkas. *19th-Century Music*, 35(3):224–237, 2012.

C. L. Krumhansl. *Cognitive Foundations of Musical Pitch*. Oxford University Press., 1990.

C. L. Krumhansl. Music psychology and music theory: Problems and prospects. *Music Theory Spectrum*, 17:53–90, 1995.

M. Laurson, M. Kuuskankare, and K. Kuitunen. The visualisation of computer-assisted music analysis information in PWGL. *Journal of New Music Research*, 37(1):61–76, March 2008.

D. Ledbetter. *Bach's Well-tempered Clavier: The 48 Preludes and Fugues*. Yale University Press, New Haven and London, 2002.

D. Ledbetter and H. Ferguson. Prelude. `http://www.oxfordmusiconline.com`, 2012.

F. Lerdahl. Atonal prolongational structure. *Contemporary Music Review*, 4:65–87, Summer 1989.

F. Lerdahl. *Tonal Pitch Space*. Oxford University Press, 2001.

F. Lerdahl and R. Jackendoff. *A Generative Theory of Tonal Music*. The Massachusetts Institute of Technology, 1983.

D. Lewin. Some ideas about voice-leading between pcsets. *Journal of Music Theory*, 42(1):15–72, 1998.

D. Lewin. Special cases of the interval function between pitch-class sets X and Y. *Journal of Music Theory*, 45(1):1–29, 2001.

M. Lindley. Temperaments. `http://www.oxfordmusiconline.com`, 2009.

R. Littlefield and D. Neumeyer. Rewriting Schenker: Narrative-history-ideology. *Music Theory Spectrum*, 14(1):38–65, Spring 1992.

Y. W. Liu. Modelling music as Markov chains: Composer identification, 2002.

J. London. Rhythm, fundamental concepts and terminology. `http://www.oxfordmusiconline.com`, 2007.

S. T. Madsen. Evolving Palestrinian counterpoint with an evolutionary algorithm. In *The 18th International FLAIRS Conference*, 2005.

S. T. Madsen and G. Widmer. Separating voices in MIDI. In *Proceedings of the 7th International Conference on Music Information Retrieval (ISMIR 2006)*, 2006.

A. Mann. *Steps to Parnassus: The Study of Counterpoint*. J. M. Dent & Sons Ltd, 1943.

A. Mann. *The Study of Counterpoint: from Johann Joseph Fux's Gradus and Parnassum*. The Norton Library, 1971.

A. Marsden. Automatic derivation of musical structure: A tool for research on Schenkerian analysis. In *Proceedings of the International Conference on Music Information Retrieval (ISMIR)*, pages 55–58, 2007.

A. Marsden. Schenkerian analysis by computer: a proof of concept. *Journal of New Music Research*, 39:269–289, 2010.

A. Marsden. Software for schenkerian analysis. In *Proceedings of the International Computer Music Conference (ICMC)*, pages 673–676, 2011.

Alan Marsden. Modelling the perception of musical voices: a case study in rule-based systems. *Computer Representations and Models in Music*, pages 239–263, 1992.

M. Mauch and S. Dixon. Simultaneous estimation of chords and musical context from audio. *IEEE Transactions on Audio, Speech, and Language Processing*, 18(6):1280–1289, 2010.

M. Mauch, S. Dixon, C. Harte, M. Casey, and B. Field. Discovering chord idioms through Beatles and Real Book songs. In *Proceedings of ISMIR 2007 Vienna, Austria.*, pages 255–258, 2007.

M. Mauch, K. Noland, and S. Dixon. Using musical structure to enhance automatic chord transcription. In *Proceedings of the 10th International Conference on Music Information Retrieval (ISMIR 2009)*, pages 669–674, 2009.

H. J. Maxwell. An expert system for harmonic analysis of tonal music. In Mira Balaban, Kermal Ebcio, and Otto Laske, editors, *Understanding Music with AI: Perspectives on Music Cognition*, chapter 13, pages 334–353. MIT Press, Cambridge, MA, USA, 1992.

M. McFarland. Transpositional combination and aggregate formation in Debussy. *Music Theory Spectrum*, 27(2):187–220, Fall 2005.

A. M. McL and Professor Dent. Voice-leading. *Music and Letters*, 31(2): 184–185, 1950. URL `http://www.jstor.org/stable/729139`.

M. McVicar, N. Yizhao, R. Santos-Rodriguez, and T. De Bie. Using online chord databases to enhance chord recognition. *Journal of New Music Research*, 40(2):139–152, 2011.

L. Mearns and S. Dixon. An empirical approach to musical style. In *Proceedings of the 3rd International Conference of Students of Systematic Musicology, Cambridge, UK*, 2010.

L. Mearns, D. Tidhar, and S. Dixon. Characterisation of composer style using high-level musical features. In *MML '10 Proceedings of 3rd International Workshop on Machine Learning and Music*, pages 37–40, 2010.

L. Mearns, E. Benetos, and S. Dixon. Automatically detecting key modulations in J. S. Bach chorale recordings. In *Sound and Music Computing, Padova*, 2011.

W. Mellors. *The Sonata Principle*. Barrie and Jenkins, 1957.

L. B. Meyer. *Explaining Music; Essays and Explorations*. The University of Chicago Press, 1973.

J. Montagu. Temperament. `http://www.oxfordmusiconline.com`, 2012.

E. Morales and R. Morales. Learning musical rules. In *IJCAI-95 International Workshop on Artificial Intelligence and Music*, 1995.

R. D. Morris. Voice-leading spaces. *Music Theory Spectrum*, 20(2):175–208, 1998.

E. Narmour. *The Analysis and Cognition of Melodic Complexity: the implication realization model*. The University of Chicago Press, 1992.

C. Neidhöfer. A theory of harmony and voice leading for the music of Olivier Messiaen. *Music Theory Spectrum*, 27(1):1–34, Spring 2005.

K. Noland. *Computational Tonality Estimation: Signal Processing and Hidden Markov Models*. PhD thesis, Queen Mary, University of London, 2009.

C. V. Palisca. *Baroque Music*. Prentice Hall, 1981.

C. V. Palisca and B. C. J. Moore. Consonance. `http://www.oxfordmusiconline.com`, 2012.

T. Pankhurst. *Schenker Guide: A Brief Handbook and Website for Schenkerian Analysis*. Routledge, 2008.

B. Pardo and W.P. Birmingham. Algorithms for chordal analysis. *Computer Music Journal*, 26(2):22–49, Summer 2002.

R. S. Parks. Voice leading and chromatic harmony in the music of Chopin. *Journal of Music Theory*, 20(2):189–214, 1976.

R. Parncutt. *Harmony: a psychoacoustical approach.* Springer-Verlag, 1989.

W. Piston. *Harmony*. W. W. Norton and Company, 1983.

D. Ponsford, G. Wiggins, and C.. Mellish. Statistical learning of harmonic movement. *Journal of New Music Research*, 28(3):150–177, Fall 1999.

L. Rabiner. A tutorial on hidden Markov models and selected applications in speech recognition. *Proceedings of the IEEE*, 1989.

J. P. Rameau. *Treatise on Harmony (1722)*. Courier Dover Publications, 1971 (Republished).

C. Raphael and J. Stoddard. Functional harmonic analysis using probabilistic models. *Computer Music Journal*, 28(3):45–52, Fall 1984.

H. Riemann. *J. S. Bach's Wohltemperirtes Clavier*. Augener Ltd, London, translated by J. S. Shedlock edition, 1890.

L. A. Roberts and M. L. Shaw. Perceived structure of triads. *Music Perception*, 2:95–124, 1984.

M. Rohrmeier. Modelling dynamics of key induction in harmony progressions. In *Proceedings of Sound and Music Computing Conference*, 2007.

M. Rohrmeier. Towards a generative syntax of tonal harmony. *Journal of Mathematics and Music*, 5(1):35–53, 2011.

M. Rohrmeier and I. Cross. Statistical properties of harmony in Bach's chorales. In *Proceedings of the 10th International Conference on Music Perception and Cognition*, pages 619–627, 2008.

C. Rosen. *The Classical Style*. Faber and Faber, 1971.

M. Ryynänen. *Automatic Transcription of Pitch Content in Music and Selected Applications*. PhD thesis, Tampere University of Technology, 2008.

F. Salzer. *Structural Hearing; Tonal Coherence in Music*. Dover Publications Inc, New York., 1982.

C. Schachter. Analysis by key: Another look at modulation. *Music Analysis*, 6(3):289–318, October 1987.

H. Schenker. *Free Composition*. Longman, 1979.

H. Schenker and F. (Ed) Salzter. Five graphic music analyses. Dover Publications Inc, New York, 1969.

A. Schönberg. *Theory of Harmony*. University of California Press, third edition, 1922.

D. Schulenberg. *The Keyboard Music of J. S. Bach*. Routledge, 2nd edition, 2006.

H. Siegel. Looking at the urlinie. In David Gagn and Poundie Burstein, editors, *Structure and Meaning in Tonal Music: A Festschrift for Carl Schachter*, chapter 7, pages 79–99. Pendragon Press, Hillsdale, NY, USA, 2006.

J. B. L. Smith, J. A. Burgoyne, D. De Roure, and S. J. Downie. Design and creation of a large-scale database of structural annotations. In *Proceedings of the 12th International Conference on Music Information Retrieval (ISMIR 2011)*, pages 669–674, 2011.

S. Strunk. Harmony (i). In B. Kernfeld, editor, *The New Grove Dictionary of Jazz*. Oxford University Press, 1988.

D. Temperley. *The Cognition of Basic Musical Structures*. MIT Press, Cambridge, MA, USA., 2001.

D. Temperley. *Music and Probability*. The Massachusetts Institute of Technology, 2007.

Y. Tomita. *J. S. Bach's Well Tempered Clavier, Book II: A Study of its Aim, Historical Significance, and Compiling Process*. PhD thesis, Leeds University, 1990.

Y. Tomita. 'most ingenious, most learned, and yet practicable work': The english reception of bach's well-tempered clavier in the first half of the nineteenth century seen through the editions published in london. In T. Ellsworth and S. Wollenberg, editors, *The Piano in Nineteenth-Century British Culture: Essays on Instruments, Performers and Repertoire*, pages 33–67. Aldershot: Ashgate, 2007a.

Y. Tomita. Anna Magdalena as Bach's copyist. *Understanding Bach*, 2: 59–76, 2007b.

D. F. (Ed) Tovey and H. (Fingering) Samuel, editors. *J. S. Bach Forty Eight Preludes and Fugues, Book One*. The Associated Board of the Royal Schools of Music, 1924.

A. Volk and A. Honigh. Mathematical and computational approaches to music: challenges in an interdisciplinary enterprise. *Journal of Mathematics and Music: Mathematical and Computational Approaches to Music Theory, Analysis, Composition and Performance*, 6(2):73–81, 2012.

E. West Marvin and A. Brinkman. The effect of modulation and formal manipulation on perception of tonic closure by expert listeners. *Music Perception*, 16(4):389–407, 1999.

P. Wishart. *Harmony : a study of the practice of the great masters*. Hutchinson, London, 1956.

J. O. Young. Key, temperament and musical expression. *The Journal of Aesthetics and Art Criticism*, 49(3):235–242, Summer 1991.